

On a wing and a prayer

This issue's focus on avian flu highlights progress and incoherence in the world's response to a potential human pandemic. But the threat is enormous, and some priorities are clear enough.

Millions of people killed in highly developed countries within months. Tens of millions worldwide. The global economy in tatters. A Hollywood fantasy? No — it's now a plausible scenario. The first act, the spread of avian flu to, and probably between, humans, has already started across Asia. Unless the international community now moves decisively to mitigate this pandemic threat, we will in all probability pay heavily within a few years. Then, hard questions will be asked as to why we were not prepared.

Sceptics abound, convinced that talk of a pandemic must be scare-mongering, or scientists crying wolf. Surely with support care, drugs and vaccines, at least the rich world can easily stand up to a flu virus? After all, this is 2005, not 1918, when a flu pandemic killed up to 50 million people worldwide. But while the science and medicine of flu have advanced substantially, our ability to mount an effective public-health response has made remarkably little progress over the decades, and the potential for panic is, if anything, greater given the impact of television and the Internet.

In the 1918 pandemic, no one had immunity to a new subtype of the influenza virus. The maths of epidemiology says that pandemics are like fault lines: they inevitably give. But unlike earthquakes, pandemics tend to give warning signs, and all the alerts from Asia are now flashing red. Will it be the 'big one'? No one can say with certainty, but the H5N1 flu strain now circulating widely in Asia, and several of its

cousins, are ones to which we humans have no immunity. Accordingly, the world now needs to develop defences for the worst-case scenario. How prepared are we?

Extinguishing avian flu in poultry and pigs, the melting-pot from which a pandemic strain would probably emerge, is the job of national agriculture and veterinary departments, the United Nations' Food and Agriculture Organization, and the World Organisation for Animal Health (OIE). The public-health aspects are the responsibility of health departments and the World Health Organization (WHO). This international coalition is shaky and far from united or sure in its purpose. Its efforts are grossly underfunded, and undermined at every turn by conflicts between global public health, sovereignty and the stakes of trade and economics.

If the next pandemic were to arise five years from now, there would have been breathing space to stimulate our drug and vaccine industries to limit the damage it would cause. But that requires urgent action now. As matters stand, a vaccine against a pandemic flu would not be ready until at least six months after a pandemic

starts. Too late: by then the worst of the pandemic would already have happened.

A vaccine that can be produced more quickly demands a research effort akin to that for a strategic military weapon, not business as usual. We also need to be able to produce enough of such a vaccine to cope with the surge in demand during a pandemic. At present, the entire world production capacity can produce only enough doses for 450 million people. To stimulate an increase in capacity, we need health policies that boost demand for existing flu vaccines in ordinary years. The same goes for antiviral drugs.

But the worst-case scenario is that a pandemic starts within two years. We would have no vaccine and few drugs, and we would be dependent on governments and the WHO to try to extinguish the first outbreaks at source. That's why the first priority must be to prevent a pandemic emerging in the first place, by extinguishing the disease in animals.

Time for action

Unfortunately, the current situation does not bode well for the abilities of governments and international agencies to cope with this challenge. We should be monitoring in almost real time the genetic changes in the avian and human viruses that could herald the emergence of a pandemic strain, for example. But there is no international funding to help affected countries build decent and sustained surveillance programmes. And while outside researchers want data from affected countries, they aren't engaging enough in the meaningful collaboration needed to build trust and open sharing. The international community is not offering incentives, such as drugs for the Asian countries that would be in the front line of a pandemic. Combine this with the fact that countries are reluctant to share the few data they have because their analysis could affect their trade and economies, and the current mess in surveillance is hardly surprising.

Each human case that occurs in Asia is potentially a global threat. The international virology community needs to be permanently there, on the ground. We need to diagnose cases swiftly, and treat the patients and all their contacts immediately with antiviral drugs to try to kill the pandemic at source. To understand the genet-

ics, and link this to the epidemiology and pathology of the virus, we need immediate sharing of all virus samples and data. None of this is happening adequately. National governments' performance is half-hearted, incomplete and far too slow. International organizations are working with their hands tied behind their backs, for bureaucratic and diplomatic reasons. In short, the level

"While the science and medicine of flu have advanced substantially, our ability to mount an effective public-health response has made remarkably little progress."

"The level of current efforts is not commensurate with the scale of the threat we face."



of current efforts is not commensurate with the threat we face.

This week, we focus on the issues in depth (see pages 390 and 399), and are providing a freely available, comprehensive collection of previous articles on the topic, not only from *Nature* but also from all other relevant Nature publications (see www.nature.com/nature/focus/avianflu/index.html). *Nature* is also engaged in a collaboration with two other organizations. The journal *Foreign Affairs* will be publishing a survey of the policy aspects of avian flu and other

pandemics in its next issue, to be published in late June. And, with both journals' involvement, the Royal Institution World Science Assembly is organizing a high-level international meeting chaired by Rita Colwell, former director of the US National Science Foundation, that is intended to bridge the gaps between science and policy.

Above all, greater top-level political oversight of the campaign is needed. The time for diplomacy and denial is over. It is time for advocacy and action. ■

Europe's constitution

Referenda next week could derail the European project — with negative consequences for science.

The people of France and the Netherlands will vote next week on whether their respective governments should ratify the proposed European constitution. Despite the traditional roles of both nations as stalwart supporters of greater European unity, their leaders have failed to generate much popular enthusiasm for the document, and both votes are expected to be close.

Research and innovation are critical to Europe's future, but have failed to emerge as an issue in the referendum campaigns. That's a shame, because their successful pursuit could hinge on the outcome of these votes.

The handful of pages in the lengthy constitution document that deal directly with research read blandly, and have inspired little enthusiasm in the scientific community. But there are aspects of the constitution that would herald significant changes in the science policy of the European Union (EU). Previous treaties, for example, have given the EU a remit to support research only as a means of bolstering industrial competitiveness. The constitution would authorize the EU to support science for its own sake.

The constitution would also tie up various loose ends in science

policy. The European Research Council, for example, which is being established to support curiosity-driven research, has no legal basis in existing EU statutes, and might be contested by any member state that chose to oppose it. The constitution brings this badly needed new agency safely within the legal remit of the EU.

And the document gives the European Commission powers to remove "legal and fiscal obstacles" to scientific cooperation across borders. It also embraces the right to conduct scientific research "free of constraint", and upholds academic freedom in universities. These components would be steps towards a more open and democratic research system.

Additionally, the constitution is the only instrument on the table that will allow the EU to develop politically, by removing the veto powers of individual states on the Council of Ministers, and by further strengthening the European Parliament. If these reforms succeed, they will help to confer much-needed legitimacy on EU institutions, and better enable the union to represent its 450 million people on the world stage.

Progress on this has been slow and cumbersome, and the verbosity of the proposed constitution reflects this challenge. Rejecting it will change nothing for scientists who find the EU to be remote and bureaucratic. Accepting it will, at least, open up opportunities for those who want to strengthen European science. ■

"The successful pursuit of research and innovation in Europe could hinge on the outcome of these votes."

Chemical biology is here

Nature and its new sibling *Nature Chemical Biology* reflect an important multidisciplinary trend.

Chemical biology is a recent addition to the scientific lexicon, and although its origin involves the use of small molecules to perturb and study biological function, it has recently grown to encompass a wide array of science at the interface between chemistry and biology. Like other multidisciplinary fields, chemical biology thrives because chemists and biologists have unique perspectives and skills that complement each other. For this reason, these collaborative efforts may be able to unravel complex biological problems.

The importance of this growing field can be seen in recent policy initiatives. In 2003, the creation of the US National Institutes of Health Roadmap (<http://nihroadmap.nih.gov>) led, for example, to the establishment of chemical-genomics screening centres and

PubChem (<http://pubchem.ncbi.nlm.nih.gov>), a cheminformatics database that is the small-molecule equivalent of PubMed. In addition, several of the recently appointed Howard Hughes Medical Institute investigators (www.hhmi.org/news/032105_list.html) have a significant chemical component to their research.

A year ago, *Nature* boosted its editorial resources specifically to respond to this trend. And this month our publishers have gone a major step further and launched *Nature Chemical Biology* (see www.nature.com/nchembio). As with all Nature journals, the new one represents a desire to meet the needs of a community without in any way reducing the commitment of *Nature* itself to publish high-quality papers in the field. Our aim is that *Nature Chemical Biology* will rapidly become the home of the strongest research for chemical biologists.

Chemical biology is often the lens that allows the biological community to see what chemists are capable of doing. It is our intention, in both *Nature* and *Nature Chemical Biology*, to illuminate the strengths and needs of these two communities and to stimulate new collaborations and scientific insights. ■

RESEARCH HIGHLIGHTS

Damp feeling

EMBO J. doi:10.1038/sj.emboj.7600668 (2005)

The bacterial flagellum has revealed an unexpected talent: it can sense wetness. The signal it produces is relayed to genes that control the injection of virulence factors, such as toxins, into the bacterium's host cell.

The flagellum is a whip-like protein with a well-known role in helping bacteria to swim. Its extra function was discovered by Rasika Harshey of the University of Texas at Austin and her colleagues, who observed that *Salmonella typhimurium* (pictured) failed to sprout full-length flagella in dry environments.

Such environments cause problems with the secretion of the filament protein that makes up the flagellum. This causes a cascade of molecular events inside the cell that ultimately affects gene regulation.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

PARTICLE PHYSICS**Lattice work**

Phys. Rev. Lett. **94**, 172001 (2005)

A promising method of calculating the interactions between fundamental particles has passed an important test. It successfully predicts the mass of the charmed B meson.

The method, unveiled early last year, is an improved version of lattice quantum chromodynamics. Researchers can now make precise predictions of particle properties using less computing power.

An international team of theorists used the method to predict the mass of the two-quark charmed B meson to be $6,304 \pm 20$ mega-electronvolts. This tallies with a measurement of the particle's mass made at Fermilab's Tevatron accelerator in Batavia, Illinois, which is being prepared for publication.

CHEMISTRY**Homeric electrons**

Angew. Chem. doi: 10.1002/anie.200500541 (2005)

A cunning metal-oxide cluster molecule acts as a Trojan Horse by carrying electrons into reactions, reports a group led by Leroy Cronin from the University of Glasgow, UK.

The colourless tungsten complex $[W_{18}O_{56}(SO_3)_2(H_2O)_2]^{8-}$ contains two pyramid-shaped sulphite groups that usually act as passive structural components. But when the complex is heated to about 400 °C, the sulphites reconfigure to form extra bonds to oxygen atoms within the cluster. This turns them into tetrahedral sulphate groups, and releases two electrons.

It is the first example of a polyoxometallate

compound containing embedded electron-releasing groups. Such molecules may have useful catalytic or electrochemical properties.

PLANT DEVELOPMENT**Surviving the split**

Plant Cell doi:10.1105/tpc.105.032185 (2005)

Short pieces of RNA that play a key role in plant development have been doing their job since before the evolution of flowers, say Michael Axtell and David Bartel of the Whitehead Institute in Massachusetts.

They used microarrays to measure the accumulation of 63 microRNAs and other RNAs that silence genes in *Arabidopsis thaliana*, a plant of the mustard family. They applied the same probes to species of wheat, pine (*Pinus resinosa*, pictured), fern and moss.

Despite the fact that these plants diverged hundreds of millions of years ago, they contain some of the same microRNAs acting on similar target genes.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

CELL BIOLOGY**Go one better**

Nature Methods doi:10.1038/nmeth764 (2005)

Enzymes that play a diverse range of roles in cellular function can sometimes be studied by creating a single mutation in their sequence. And when one mutation doesn't work, two might, says a team headed by Kevan Shokat of the University of California, San Francisco.

Five years ago, Shokat's group discovered a way to target specific protein kinases by changing their structure: substrates or inhibitors target the kinase of interest after a simple mutation removes a 'gatekeeper' amino acid from in front of one of the binding sites. Unfortunately, the mutation can destroy an enzyme's activity, thwarting attempts to study its function. Now the team has found that a second mutation can rescue activity.

MATERIALS**No flat batteries**

Adv. Mater. **17**, 1230-1233 (2005)

Betavoltaic batteries create current from the electrons (β -particles) that are produced by a radioisotope when its neutrons decay to protons. They tend to be inefficient, but a design that may improve their performance has emerged from Philippe Fauchet's team at the University of Rochester, New York.

Conventional designs pass tritium gas over flat electron collectors. Fauchet's group fill a porous silicon block with the radioactive gas instead. This has a greater surface area for electron collection, boosting the current by a factor of ten. The team estimates that almost

every β -particle in the battery contributes to the current. Using radioisotopes with long half-lives could produce low-power batteries that last for decades — ideal for satellites.

CHEMICAL BIOLOGY

Ironing out bugs

Nature Chem. Biol. **1**, 29–32 (2005)

The bacteria responsible for tuberculosis and plague both need iron to become virulent in humans, and both produce iron-sequestering compounds, known as siderophores. So drugs that block siderophore synthesis could provide a line for antibiotic attack.

Researchers from Cornell University and the Memorial Sloan-Kettering Cancer Center in New York have devised a molecule that binds to and inhibits enzymes involved in siderophore synthesis. The compound successfully reduces the growth of both *Mycobacterium tuberculosis* and *Yersinia pestis* under iron-poor *in vitro* conditions.

CANCER

To catch a kinase

Nature Genet. doi:10.1038/ng1571 (2005)

In order to divide and spread, many cancers rely on faults in enzymes called protein kinases, which means mutant kinases are potential drug targets. Humans have 518 known kinase genes, and a team headed by Mike Stratton of the UK Sanger Institute has screened the DNA sequences of all these genes in 25 cases of breast cancer.

The team found the number of non-inherited mutations in the cancer cells ranged from none to dozens. The pattern of changes in some cases suggests there is an underlying mutation mechanism that has not been seen before.

NONLINEAR PHYSICS

Black magic

Phys. Rev. Lett. **94**, 184503 (2005)

The bizarre black disc pictured right is a magnetic fluid deformed by solitons — localized packets of energy that have a stable shape.

Each spike is a soliton, about a centimetre high, that rises from the surface of a suspension of iron oxide particles. Created by Reinhard Richter at Germany's Bayreuth University and Igor Barashenkov at the University of Cape Town in South Africa, the structures are held aloft in the ferrofluid by a constant magnetic field.

The solitons seem to keep their shape because of the nonlinearity that the spike creates in the surrounding magnetic field. Similar mechanisms have previously been observed only in one-dimensional solitons. Other surface solitons, which are confined in two dimensions, need energy to sustain their structure.

QUANTUM DOTS

Light boxes

Science **308**, 1158–1161 (2005)

Put a nanoscale blob of semiconductor inside a cavity that contains light and it is possible to probe the fundamentals of quantum mechanics by looking at how this 'quantum dot' and the light interact. Sounds straightforward, but research has been hindered by construction problems — it's hard to put the dot in the cavity, and difficult to design the cavity to trap the right wavelength of light.

An encouraging approach for building these systems in semiconductors comes from researchers at the University of California,



APS

Santa Barbara, and the Swiss Federal Institute of Technology in Zurich. They grew a dot in a sandwich of layers, then drilled holes around it to define a 'photonic crystal' that acts as cavity walls. They tuned the wavelength by varying the size of the holes.

DEVELOPMENTAL BIOLOGY

The right slant

Cell **121**, 633–644 (2005)

One of the biggest questions in developmental biology is how embryos that begin as uniform balls of cells end up asymmetric. Experiments in mice provided a clue when researchers discovered that hair-like cilia protruding from embryonic cells in mice rotate, somehow setting up a flow in the surrounding fluid that defines the left–right axis.

A team led by Nobutaka Hirokawa of the University of Tokyo supplies another piece of the puzzle by showing exactly how the cilia's movement sets up the directional flow. The researchers found that the cilia rotate around an axis tilted 40° backwards. The same rotation was observed in rabbit and medakafish embryos, suggesting that the mechanism, previously studied only in mice, defines asymmetry in other vertebrates.

JOURNAL CLUB

Thomas F. Stocker
University of Bern, Switzerland

A climate physicist describes plans to monitor the disaster scenario that Hollywood turned into a blockbuster film.

Soon after the opening credits of *The Day After Tomorrow*, a couple of buoys floating in the Atlantic Ocean register plummeting water temperature. The film gets into full swing as the Gulf Stream shuts down, which takes hours. Days later, ice blankets northern Europe and the United States.

Events would not unfold as fast as the film suggests, but widespread freshening of the Atlantic caused by more rain and melting ice could weaken the north-flowing part of the Gulf Stream. If the current stopped completely, Europe would suffer without the warmth it brings.

Some researchers entertain 'dreamlike scenarios' of such climate turmoil, but I am reassured that others are asking relevant questions: what signs would there be of an imminent shutdown, and where in the Atlantic

might we be able to detect them?

The most precise recipe comes from the *Journal of Marine Research* (J. Baehr *et al.* **62**, 283–312; 2004). The authors use a computer model to show that temperature and salinity data from moorings along a latitude of 26° N, combined with measurements of surface winds and estimates of current strength off Florida, could give the strength of the Atlantic current system to within 10% — and predict its fate.

"We still know very little about ocean currents."

I think investment in such an array would be worth while, even if today's widespread freshening turns out not to be a harbinger of disaster. In 1984, my predecessor at Bern's Physics Institute, Hans Oeschger, proposed that flips in circulation patterns could trigger abrupt climate change, but we still know very little about ocean currents. An Atlantic array would allow modelling groups such as mine to improve our predictions by quantifying the natural variability of the ocean circulation for the first time.

NEWS

Bird flu spreads among Java's pigs

TOKYO

Concerns over the presence of a dangerous strain of avian flu virus in Indonesia's pigs are growing, as government tests confirm the existence of infection. In some areas, the H5N1 virus could be infecting up to half of the pig population, without causing any signs of disease.

The initial discovery was made earlier this year by an independent researcher working outside national and international surveillance systems. Chairul Nidom, a virologist at Airlangga University's tropical-disease centre in Surabaya, Java, found the H5N1 virus in five of ten pigs tested from Banten in western Java.

The presence of the virus in pigs is a particular worry because the animals can harbour both bird and human flu viruses, and act as a 'mixing vessel' for the emergence of a strain of avian flu that can easily infect humans. There are now signs that the virus could be spreading unchecked through the pig population.

Nidom says that the pigs he tested showed no signs of illness, and the only reason he tested them was that they were kept near a chicken farm that was struck by avian flu last year. *Nature* has discovered that a government survey has since found similar results in the same region.

The virus was not found in 150 pigs tested from outside the area. Although the government says it has stepped up the surveillance of pigs in its seven satellite laboratories, it may fail

to spot any spread of the virus because resources are short. "It's a big country," says Tri Satya Putri Naipospos, director of animal health at Indonesia's agriculture ministry. "If you want to commit to eradicating a disease, you need more money."

Nidom is also frustrated by a lack of resources. He says he has samples from another

90 pigs in Banten, but cannot afford to test them or to expand his survey to other areas.

Some health officials in Asia fear the presence of avian flu in pigs even more than in chickens or ducks. "I think pigs pose a much greater threat of spreading the disease to humans than poultry," says Nidom.

The virus was found in pigs in China in

R. HUIBERS/PANOS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Mixing vessel: pigs can host both human and avian flu viruses, which could encourage a dangerous merger.

Ecologist's tenure hailed as triumph for academic freedom

SAN DIEGO

In a closely watched test case for academic freedom, the University of California, Berkeley, has granted tenure to an outspoken ecologist.

Ignacio Chapela, who once led faculty objections to Berkeley's research pact with Novartis, was told on 17 May that his five-year quest for tenure had finally succeeded. "It is just amazing," Chapela told *Nature*. "This is like a new start in academia."

In the United States and beyond, the Chapela tenure fight has symbolized the conflict between academics' freedom to challenge policies, and the relationship between public universities and industry. The case has taken on special significance in recent years as professors at other US

universities have come under attack for their views, with leading politicians calling for reduced tenure protections.

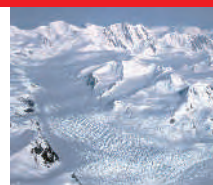
The Berkeley campus has become known for its battles over free speech, and as word about the tenure approval spread, there was elation among supporters of the Mexican researcher. "This has restored my faith in the institution," says Wayne Getz, an environmental scientist who championed Chapela's cause.

In 1998, the College of Natural Resources at Berkeley accepted \$25 million from the Switzerland-based Novartis (now Syngenta). Shortly afterwards, Chapela raised questions about the impact of the five-year deal on research and teaching. The issue bitterly divided the campus into those

who supported and opposed the deal.

Chapela's quest for tenure was complicated by a controversial article he wrote for *Nature* about the flow of transgenes into wild maize varieties in Mexico (I. H. Chapela and D. Quist *Nature* 414, 541–543; 2001) — parts of which were later called into question.

Faculty committees supported Chapela's application for tenure, but in 2003 it was denied by the chancellor at the time, Robert Berdahl, who acted on advice from the final review committee. Chapela appealed to the faculty senate, alleging that the process had been corrupted by pro-industry faculty members. An academic senate inquiry last year decided that there were irregularities, setting the stage for another review by a newly constituted, high-level committee.

**WEIGHTY TOPIC**

The East Antarctic ice sheet is getting thicker from increased snowfall, slowing sea-level rise
www.nature.com/news

BAS

2001 and in 2003 (see *Nature* **430**, 955; 2004). The country stepped up its surveillance, and two surveys in 2004 found that all 8,457 samples tested were free of H5N1.

Nidom's discovery of H5N1 in pigs is a wake-up call for the Indonesian government. He says that when he first alerted the government to his findings in February, there was no reaction. "I don't know why they are so passive," he says. Nidom took his findings to a local newspaper, the Jakarta-based *Kompas*, which ran the story on 9 April. The news spread to international media earlier this month.

The government responded to the media attention by carrying out its own survey, and found H5N1 in three out of eight pigs it tested in Banten, Naipospos told *Nature*. Like those tested by Nidom, the pigs showed no outward signs of disease.

Despite this worrying result, communication has faltered between Indonesia and the international organizations charged with monitoring animal health, such as the United Nations' Food and Agriculture Organization (FAO) and the World Organisation for Animal Health (OIE). When interviewed by *Nature* last week, the OIE's regional representative for the Asia-Pacific region still referred to the presence of H5N1 in pigs as "a rumour".

The FAO and OIE cannot act until they have received official government reports, says Carolyn Benigno, animal-health officer at the FAO's regional headquarters in Bangkok. She hadn't heard of Nidom's work until *Nature* contacted her last week. However, Naipospos complains that although she is preparing an official report for the FAO, she cannot fast-

FLU BULLETINS

The past week has seen the release of worrying data on the risk of a human pandemic, alongside almost daily news of further cases of avian flu.

VIRUS EVOLVING On 18 May, the World Health Organization (WHO) confirmed that the epidemiology of human outbreaks of avian flu in northern Vietnam this year is different from that elsewhere, and could be consistent with human-to-human transmission. There are not enough data to make a direct link, but the WHO has also found genetic changes in strains from northern Vietnam that could be consistent with the virus evolving to be more infectious.

Sequencing data of isolates from several countries also suggest more rapid evolution. Last year, all H5N1 samples were broadly similar. But this

year, viruses from northern Vietnam and Thailand form a separate cluster from those isolated from southern Vietnam and Cambodia.

TAMIFLU RESISTANCE FOUND H5N1 isolated from a patient in northern Vietnam was found to have partial resistance to oseltamivir (Tamiflu), the antiviral drug that countries are rushing to stockpile (see page 407). Tamiflu is the drug of choice for treating H5N1. More monitoring of resistance is urgently needed, the WHO warned in a report on 18 May.

CASES INCREASE On 19 May, the WHO reported a rise in cases in the current outbreak of human H5N1 infections — up from 89 cases and 52 deaths on 4 May to 97 cases and 53 deaths, in Vietnam, Thailand and Cambodia. Last

weekend, the Vietnamese media announced three further cases, one of whom has died.

VIRUS REACHES MIGRATORY BIRDS On 21 May, China announced that H5N1 was responsible for the deaths of dozens of migratory bar-headed geese in the western province of Qinghai. China has sealed off areas around Qinghai lake, which is a migratory stop for hundreds of thousands of birds, and has sent 3 million doses of poultry vaccine to the region.

CONFUSION IN INDONESIA Out of 79 Indonesian poultry workers tested for H5N1, one sample has been confirmed as positive. As no symptoms were reported, the government had hoped to check another sample, but hasn't been able to find the man concerned.

Declan Butler

track it because the FAO and the OIE do not classify the case as an emergency. "This is not an outbreak, it's a finding," she says, because the pigs are not ill or dying. As *Nature* went to press, the Indonesian government was preparing to send a report on the matter to the OIE.

Nidom says he would like to expand his pro-

ject, and to sample pigs from eastern Java. But he is not counting on being given the resources to do so. This is his second run-in with the government — in 2003 he caused a stir by releasing data showing that mass deaths of chickens at the time were caused by H5N1. ■

David Cyranoski

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Last month, that committee advised Robert Birgeneau, the current chancellor, to grant tenure, which he did last week. Birgeneau was unavailable for comment, but university spokesman George Strait insists the tenure process was normal, if

unusually protracted. "This shows the process works," he says. Chapela, now ranked as associate professor, has a slightly different view. "To me, it shows the system doesn't work, unless you fight," he says.

Chapela plans to reopen his laboratory,

Ignacio Chapela has finally won tenure after a very public battle with the University of California, Berkeley.

which was shut down in 2002 when the ongoing tenure battle affected his ability to secure grants. But he is also assessing his legal options. In March, he filed a lawsuit in a state court against Berkeley, alleging discrimination, retaliation and fraud in his tenure review, which the university denies. Some of his advocates are urging him to continue the fight for compensation.

"This is the first step. The true measure of the university's commitment to academic freedom will be its willingness to fix a system rife with problems," says David Quist, Chapela's most recent doctoral student.

Chapela's anti-industry campaigns continue. He and his supporters spent every night last week cycling round the construction site of a bioengineering centre on the Berkeley campus, to draw attention to planned industrial partnerships. ■

Rex Dalton

UK panel urges animal researchers to go public

LONDON

Years of sometimes violent protests have left many British biomedical researchers afraid to talk about research involving animals. So it is with some trepidation that scientists face a call from the Nuffield Council on Bioethics to be open about what happens to animals in their labs.

Abusive letters and e-mails are regularly sent to the UK scientists who speak out in favour of animal research. In extreme cases, researchers have been physically attacked or had bombs sent to their homes. As a result, only a tiny fraction of Britain's thousands of biomedical scientists are willing to participate in discussions on animal research in the national media. "There are about 25 in the country," estimates Simon Festing, director of the RDS, a lobby group in London that supports animal research.

One consequence is public confusion about whether animal research is worthwhile, says Nuffield, the London-based think-tank, in a report published this week. And the solution, according to the report, is for researchers to get more involved in debates on the subject. It also wants the government to publish more details about the procedures used in animal experiments. About 3 million animals are used in scientific research each year in Britain, of which 80% are mice and rats.

"Openness will lead to better dialogue," says Steve Brown, an author of the report and director of the Medical Research Council's Mammalian Genetics Unit in Harwell near Oxford.

That principle is backed by critics and advocates of animal research — both sides claim

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

What are you looking at? Scientists are being advised to give details of their animal experiments.

that the more the public knows, the more it will support their arguments (see 'Straight talking').

"It will operate in our favour," says Alistair Currie, campaigns director of the British Union for the Abolition of Vivisection, also based in London. The union, one of four animal-welfare groups represented on the report's 18-member panel, says few people realize how much basic research involves the use of animals. It claims the public will be more likely to oppose research if such information is available.

"The antivivisectionists will run highly emotive campaigns. But I have faith in the process of democracy."

"The antivivisectionists will run highly emotive campaigns, but they do that anyway," counters Festing. "I have faith in the process of democracy. People can see that there is a clear link between the animals' suffering and medical advances."

However, researchers fear that releasing any extra information about their work could make them more vulnerable to attack.

A recent government decision to place online a summary of every licence issued for animal studies was praised by the Nuffield report. And the council wants the anonymous summaries, which started going up last December, to include information supplied after the experiments about, for example, the level of suffering caused and whether scientific objectives were met. But scientists fear that such details increase the risk to individual researchers.

"If you get down to the specifics of methodologies, you can pinpoint the people involved," says one senior neuroscientist, who asked not to be named.

Festing backs the summaries, but says that another way to spread information is to provide safe places for researchers to discuss their work. He suggests that universities could hold internal debates between faculties, to give researchers the confidence to speak openly.

Government commitment to protecting scientists is crucial, Festing adds. Legislation aimed at restricting activists' protests came into force on 7 April. "We need proper policing and enforcement of that law," he says. "If that goes ahead it will boost confidence." ■
Jim Giles

STRAIGHT TALKING

The procedures detailed here, taken from research papers and animal-study licences, include the type of information that scientists are being asked to discuss with the public. But will such details make people more supportive of animal research, or less?

EXPERIMENT 1

Animals used: Mice, rats, guinea-pigs, rabbits and dogs

Why? To find the doses of new drugs that are suitable for testing in humans.

What happens? The animals are given increasing doses of test drugs over a 24-hour period, by mouth, injection or a tube inserted in a blood vessel. During this time, the animals are restrained. To reduce the number of animals used, up to ten drugs may be given

to each animal. Regular samples of blood are taken and bile may be tapped under general anaesthetic. The animals are killed if they show any discomfort, or at the end of the experiment.

EXPERIMENT 2

Animals used: Rhesus monkeys

Why? To assess whether the brain chemical GDNF can prevent the death of dopamine-producing brain cells in monkeys that have been given a condition that is

similar to Parkinson's disease.

What happens? MPTP, a chemical that induces a Parkinson's-like state, is injected into the monkeys' arteries. In animals that develop full symptoms, GDNF or a control is injected into their brain, with the needle held in place for three minutes. After a week the monkeys undergo regular hand-reaching tasks for three months. They are then anaesthetized, given a positron emission tomography scan, and killed.

EXPERIMENT 3

Animals used: Mice

Why? To test whether genetically modified immune-system cells can treat cancer.

What happens? Cancer cells are injected under the skin of mice to trigger a tumour. Genetically modified immune cells are injected into the tail, and tumour growth is measured using callipers. The mice are killed if the tumour grows big enough 'to cause discomfort', or at the end of the experiment. J.G.

Korea's accelerating stem-cell work prompts calls for global ethical rules

JUNG YEON-JE, POOL/AP

A landmark paper in stem-cell research has been seized on by both sides of the cloning debate, as they seek to further their respective causes. But the advance has also led ethicists to call for broader international agreement over how such studies should be conducted.

On 19 May, a team led by Woo Suk Hwang of Seoul National University in South Korea reported that it had created 11 human embryonic stem-cell lines that are genetically matched to individual patients (W. S. Hwang *et al. Science* doi:10.1126/science.1112286; 2005).

Hwang's team is the first to report this achievement, and the work has been hailed by scientists as a major breakthrough, because it could allow researchers to follow the development of a patient's disease in a Petri dish, instead of in animal models. Eventually, the work could lead to therapies, but scientists caution that this could be years away — and might never happen.

"This is a huge advance that could help us get at a lot of diseases that remain mysterious," says George Daley, a stem-cell researcher at Harvard Medical School in Boston, Massachusetts. "I wish we were doing this work here."

But even as scientists lauded Hwang's work, others condemned it because of the technique he used to create the cell lines. The procedure, often described as 'therapeutic cloning', involves taking a skin cell from a patient and extracting its nucleus. This is then placed inside a donated human egg stripped of its own genetic material, and the egg is allowed to begin developing into an embryo. After a few days, when it has reached the stage known as a blastocyst, scientists can extract stem cells that are an exact genetic match to the patient. But critics say that creating a blastocyst and then destroying it is equivalent to creating and destroying a human life.

One of these critics is US President George Bush, who says that Hwang's research is morally troubling. "I'm very concerned about cloning," Bush told reporters on 20 May. "I worry about a world in which cloning becomes acceptable."

Bush's concern perhaps reflects the fact that US lawmakers are preparing for their first vote on a proposal to loosen rules on stem-cell research that he set on 9 August 2001. These rules prevent federal funds

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Clone ranger: Woo Suk Hwang is widely seen as being at the frontier of stem-cell research.

from being used for work on human embryonic stem-cell lines derived after that date. But revelations about weaknesses of available cell lines — and opinion polls suggesting that supporters of Bush's party may not back his policy — brought the issue to a head in the run-up to this week's vote.

Others are using Hwang's results to argue that the field should speed up outside Korea. Speaking at the Science Museum in London on 20 May, Nobel laureate James Watson lambasted Britain's "piss-poor" approach to stem-cell research. He described some UK cloning work as "nice science", but said lack of funds meant it wasn't moving fast enough. "You want to be in the same league as the Koreans, but you are a bit player," he said.

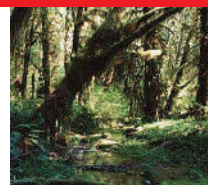
But an accelerating field would increase demand for donated eggs, which raises its own ethical problems. An international effort to coordinate stem-cell research would lend transparency to the field and ensure it proceeds in an ethical way, says Arthur Caplan, a bioethicist at the University of Pennsylvania.

South Korean Sang-yong Song, who is president of the Asian Bioethics Association, agrees. He worries that Hwang's paper will be used by some countries as an excuse to accelerate research, before consensus about how to do the work is reached. "Even in Korea there is no national consensus on promoting biotechnology," he says.

"There is a real need for international control," Caplan says. "Nations are moving," he adds, noting that people against stem-cell work shouldn't "simply stand on the sidelines pouting and saying 'you can't do this'."

Erika Check

Additional reporting by David Cyranoski



DIVERSITY DIVES

Ecosystem Assessment report says humans raise species extinctions by a thousand times

www.nature.com/news

GETTY

Zambia to wage 'scientific' war on malaria

With proper planning and commitment, malaria deaths in Africa could be cut by 75% in just three years. That's what the Bill & Melinda Gates Foundation is out to prove, and it has chosen Zambia as its battleground.

Many have been frustrated by the progress of international efforts, such as the UN-led Roll Back Malaria partnership, which in 1998 pledged to halve malaria deaths worldwide by 2010. It is nowhere near meeting its goal, and the Gates Foundation believes that planning is the missing factor.

Zambia has already declared that it will get insecticide-treated bednets to 90% of its households. It also aims to treat 60% of malaria patients with artemisinin drugs within 24 hours of diagnosis.

"The Zambian people and their leaders have shown tremendous resolve in their commitment to controlling and preventing malaria," says Richard Feachem, executive director of the Global Fund to Fight AIDS, Tuberculosis and Malaria. The Global Fund has pledged \$38 million to Zambia for malaria control over the next two years, and up to \$83 million for the next five.

The Gates initiative aims to glue these efforts together with sophisticated data-management systems and extensive scientific trials to evaluate what works and what does not. This scientific support will be funded by a \$35-million investment over the next nine years. The foundation announced the decision to aid Zambia at the World Health Assembly in Geneva on 19 May.

The world needs "a blockbuster success against malaria", says Regina Rabinovich, director of the foundation's infectious-diseases programme. She hopes the project will "encourage both donor countries and developing countries to devote greater resources to fighting malaria".

Called the Malaria Control and Evaluation Partnership in Africa (MACEPA), the project will be run by the Seattle-based non-profit organization PATH. It will partner Roll Back Malaria, the Global Fund and the World Bank.

Carlos Campbell, MACEPA programme director at PATH, says that its real-time data-management systems are "based

on the best science available". He adds that measuring the effectiveness of the efforts in Zambia is a priority, so the results can guide future large-scale efforts. Although funding

for malaria control has been stuck for years at around \$200 million annually, money is starting to flow, he says, so solid science is needed to ensure that the

influx of resources is spent well.

MACEPA will also measure the economic impact of reducing malaria deaths. The World Bank estimates that the disease cuts the economic growth of African countries by at least 1.3% annually. Brian Chituwo, Zambia's health minister, says it was the social and economic impact of malaria, which is responsible for 40% of child deaths in the country, that drove his government to make malaria control a priority.

Campbell hopes that the project will prove that investment in malaria control pays for itself through economic growth, encouraging other African countries to follow in Zambia's footsteps.

Declan Butler

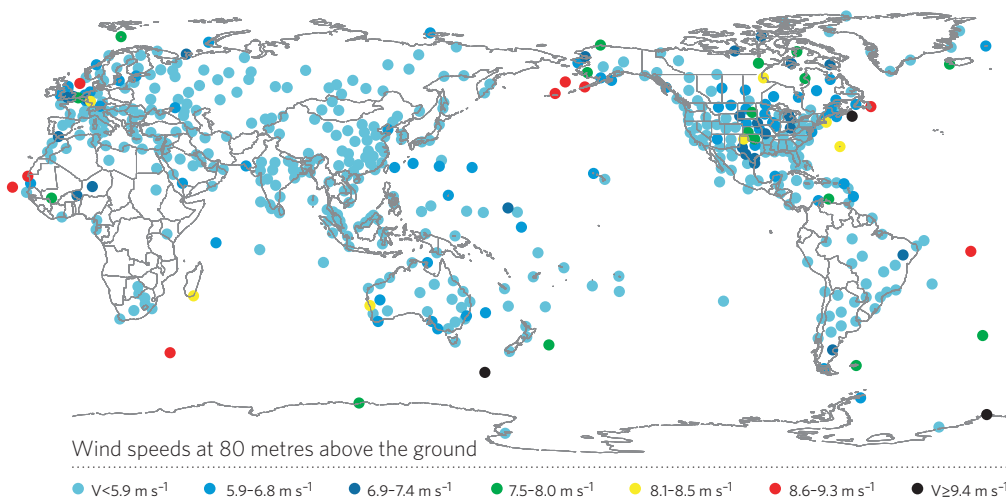
"The world needs a blockbuster success against malaria."

GRAPHIC DETAIL

Wind mappers blown away by high energy potential

To inspire us with the power of wind turbines, researchers have painstakingly assembled this map of wind speeds across the globe.

The main difficulty with judging the amount of wind available for power generation is that speeds are typically measured at stations 10 metres above the ground — but modern wind turbines are 80 metres tall. Cristina Archer and Mark Jacobson from Stanford University, California, collected wind measurements from 8,000 locations around the world. They revised existing equations to estimate what the speeds would be 80 metres up (C. L. Archer and M. Z. Jacobson *J. Geophys. Res.*, in the press). Particularly windy spots include the North Sea, the tip of



South America, Tasmania and North America's great lakes.

Any location with an average annual wind speed of 6.9 metres per second or above is considered suitable for low-cost power generation. The researchers estimate from their map that, excluding practical considerations, the world's energy needs could be met by harnessing just 20% of the planet's potential wind power.

C. ARCHER & M. JACOBSON

ON THE RECORD

“The Royal Society today is a lazy institution, resting on its historical laurels... It is little more than a shrill and superficial cheerleader for British science.”

An editorial in *The Lancet* calls for the Royal Society to revamp its mission.

“The benefits of research that kills living human embryos are purely speculative and have been hyped by researchers who are after federal funding.”

Congressman Steve King (Republican, Iowa) condemns plans for US funding of embryonic stem-cell research.

SCORECARD

**Weeds**

The rampant, invasive weed kudzu may be useful after all. An extract from the plant seems to help people reduce their alcohol consumption — a sobering thought.

**Smallpox experiments**

The World Health Organization says it might approve research on smallpox, if scientists provide more details on the work.

**Japanese girl power**

Japan finds it has a lower percentage of women in its science and technology work force than any other rich nation.

OVERHYPED

Shark cartilage

During the 1990s, shark cartilage was the ‘next big thing’ — an alternative cancer treatment that actually seemed to work, at least in some animal studies. But a paper in this week’s *Cancer* takes a thorough look at shark cartilage in human cancer patients, and it’s not pretty. The paper reports that 42 cancer patients who took the cartilage lived no longer than the 41 who went without it. In fact, those taking the cartilage felt worse. Maybe this will sink the shark idea once and for all (C. L. Loprinzi *et al.* *Cancer* doi:10.1002/cncr.21107; 2005).

We'll rain on your parade, forecasters tell rogue pundits

LONDON

Frustrated weather forecasters are fighting back against rogue companies that sell forecasts with claims of impressive accuracy, but that have no apparent scientific basis.

Private forecast providers have proliferated in Europe and the United States over the past decade as various industries demand increasingly accurate information about the weather. Insurance companies want to be warned of wet weather, for example, which can result in claims for flood damage. And retailers need shorter-term predictions to help them judge what to stock. “If there’s a heatwave coming we need barbecues, salads and lager,” says a spokeswoman for Sainsbury’s supermarkets in the United Kingdom.

Most providers are large companies that generate a high level of satisfaction among their customers. But even national meteorological offices have been accused of using unclear statistics to advertise the accuracy of their forecasts. And there is nothing to stop any individual from using freely available data, such as those released by the US National Weather Service, to produce their own weather forecasts.

Meteorologists are concerned that the rise of private forecasters could affect the reputation of their trade. They are reluctant to name specific firms for fear of legal action, but say that some forecast providers use professional-looking websites to hide the fact they are using scientifically flawed methods that produce unreliable predictions. “We have companies claiming they can predict months ahead, using methods they will not mention,” says Pascal Mailier, a meteorologist at the University of Reading, UK.

So Britain’s Royal Meteorological Society has decided to develop an accepted set of metrics with which to rate the accuracy of different forecasting firms. Mailier, who is leading the project, hopes the plan will clean up the industry. “I’m giving them the chance to prove themselves,” he says.

To come up with a fair way of rating firms, the society has asked Mailier and his colleagues to survey the methods that are available for assessing forecasts. They will present preliminary results, together with comments

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Washout: inaccurate weather forecasts can lead to huge losses for companies such as insurers.

from users and the 20 or so forecast providers in Britain, to the society’s annual conference in September. The rating system could in principle be used by an independent body to gauge the accuracy of different providers.

Finding a method that everyone can agree on, however, is likely to be less straightforward than it sounds. One way is to award points for different aspects of the forecast, such as wind speed or temperature, and add them up to give a total score. But this involves arbitrary

“We have companies claiming they can predict months ahead, using methods they will not mention.”

judgments, so the score is highly dependent on how the points are awarded.

The Met Office in Britain has traditionally rated its next-day forecasts as 85% accurate. This sounds impressive, but forecasts obtained simply by assuming that tomorrow’s weather will be the same as today’s achieve 77% (see J. E. Thornes and E. A. J. Proctor *Weather* **54**, 311–321; 1999). The Met Office now says that the figure of 85% attracts too much attention, and prefers instead to use standard statistical measures of error that focus on one variable at a time.

No single metric is likely to do the job, the project team warns, because the needs of users vary. Local governments might be interested in the accuracy of ground-temperature forecasts, for example, to decide whether roads need to be gritted in advance against

C. LUZIER/REUTERS/NEWS.COM



GAS PRIZE
NASA launches contest
to suck oxygen out of
Moon dust
www.nature.com/news

NASA

ice. Airports want to know wind speeds, and energy suppliers care about temperature, especially longer-term events such as unusually mild winters.

Meteorologists also caution that data from some rural areas are patchy, and that some aspects of weather will be hard to quantify. "The state of the sky is the most difficult," says John Thornes of the University of Birmingham, UK. Many providers forecast the cloud cover, but most measurements are generated visually by airport employees.

Despite these barriers, many users of forecasts contacted by *Nature* welcomed the society's survey plans. Major companies such as supermarkets and energy suppliers already run their own assessments of the forecasts that they buy, but say an independent set of metrics would be a useful addition.

Even forecasts from reputable providers can vary considerably in their accuracy, says Gearoid Lane of Centrica, an energy and services company based in Windsor, UK. If the forecasts are wrong, he adds, "there are big financial consequences". ■

Jim Giles

NIH hints at ethics rule change

WASHINGTON DC

Researchers who had threatened to walk away from top jobs at the US National Institutes of Health (NIH) over tough ethical rules have now changed their minds.

The scientists say they have been privately assured by NIH director Elias Zerhouni that they will not have to divest themselves of all their biomedical stock. It is unclear how such assurances will be reconciled with the rules, introduced in February as an interim measure after a series of scandals over possible conflicts of interest. It is not known whether the stock rules will be changed.

The NIH announced last week that David Schwartz, a physician at Duke University in Durham, North Carolina, would head the National Institute of Environmental Health Sciences. Earlier reports said that he was turning the job down because of rules against holding stock in certain situations.

And James Battey, who in March announced his intention to resign as head

of the National Institute on Deafness and Other Communication Disorders, changed his mind after a meeting with Zerhouni in early May. Battey was in the running to head the California Institute for Regenerative Medicine, which manages the state's stem-cell initiative.

Battey had said he was resigning because he managed a family trust fund that included biomedical stock. "I received reassurance from the NIH director that I would be able to continue my responsibility to my family and remain at the NIH," he now says.

Reviews of the regulations are ongoing at the NIH and its parent department, Health and Human Services. Many — including congressional representatives, scientific organizations, and Zerhouni himself — have said that the interim rules go too far. Permanent rules are not expected until next February. ■

Emma Marris

Grant probe leaves DNA lab without direction

The future of one of the top laboratories studying ancient DNA is in doubt after its director resigned under a cloud.

In just six years, the Henry Wellcome Ancient Biomolecules Centre at the University of Oxford, UK, has made many high-profile finds using genetic material from ancient samples.

But geneticist Alan Cooper, the centre's founding director, resigned on 22 April following a university investigation into his conduct. Oxford officials declined to discuss specifics, but issued a statement saying that: "Allegations relate solely to material included in grant applications and not to published research results or the conduct of ongoing research."

Sources say that the probe examined claims that, in a grant application to the Natural Environment Research Council, a figure was incorrectly reproduced and there were questions over certain signatures.

In an interview with *Nature*, Cooper acknowledged that there had been some problems, saying "there were a couple of errors in judgment". He is now setting up a similar lab at the University of Adelaide in South Australia.

Rift over successor to Kyoto is as wide as ever

A United Nations seminar in Bonn, Germany, last week failed to end the transatlantic impasse over climate change, especially once the Kyoto Protocol has expired in 2012.

The treaty, which came into force in February, requires industrialized countries to reduce their emissions of greenhouse gases. The European Union has called for a

NASA told to think small in hunt for asteroids

A bill sponsored by Congressman Dana Rohrabacher (Republican, California) would require NASA to extend the search for threatening asteroids and comets down to objects as small as 100 metres across. If enacted, the bill, which passed the House Science Committee last week, would authorize \$20 million a year for the next two years to get the search under way.

NASA is already on track to catalogue 90% of objects more than 1 kilometre in size by 2008. But smaller objects

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

could still cause regional or even global damage if they hit. "Any threat that would wreak havoc on our world should be studied and prevented if possible," says Rohrabacher.

A NASA committee

concluded in 2003 that the search for smaller objects could be done for less than \$400 million. But the panel's chair, Grant Stokes of the Massachusetts Institute of Technology's Lincoln Laboratory, whose LINEAR search currently accounts for

most near-Earth asteroid discoveries, says no funding agency has picked up the tab, in part because asteroid-hunting doesn't fulfil an obvious science mission. "In a sense, it's a public service," he says.

follow-up commitment that would require emissions to be cut by 15–30% relative to 1990 levels by 2020.

This goal is hardly achievable without the cooperation of the United States, which withdrew from the protocol in 2001 and is the world's largest producer of greenhouse gases. But the United States made no sign last week of going back on its brusque rejection of mandatory targets for reducing emissions.

One positive sign was a move led by several less-industrialized countries, including China and Papua New Guinea, towards more actively tackling climate change, says Karla Schoeters, director of the Climate Action Network, a European umbrella agency working on climate and energy issues.

Science adviser quits after conflict-of-interest furore

Australia's chief scientist, Robin Batterham, has resigned his government post to work full-time for the mining and energy giant Rio Tinto.

During his six-year tenure, Batterham worked part-time for the company while spending two days a week advising the government on policies including carbon dioxide emissions. Critics argued that this represented a conflict of interest.

Last year, a Senate inquiry found no evidence of improper conduct by Batterham, but recommended that the position of chief scientist should be full-time. The government has yet to decide on the terms of the next appointment. Batterham, who will have a say in the selection, says that industrial experience is crucial.

Environmental lobbyists had been angered by Batterham's dismissal of the Kyoto Protocol, which Australia refused to ratify.

Pressure eases on AIDS groups to denounce vice

There's a new twist in the ongoing debate about whether the United States can force international AIDS-fighting groups to denounce prostitution.

Earlier this month, the US Centers for Disease Control and Prevention released a document stating that agencies applying for money from the Global Fund to Fight AIDS, Tuberculosis and Malaria must take a public stand against prostitution and sex trafficking. But last week the agency changed its mind.

Randall Tobias, the Bush administration's global AIDS coordinator, has had the policy rescinded, saying he never approved it.

Public-health advocates have pointed out that prostitutes are at particularly high risk of HIV infection, and that opposing prostitution could alienate this group, hampering efforts to slow the virus's spread.

US government urged to aid foreign scientists' entry

The US government must do more to make it easier for foreign scientists to enter the country, a group of scientific and academic societies declared last week.

In a statement, the coalition of 40 societies — including the National Academy of Sciences and the Association of American Universities — warned that "despite significant recent improvements to the US visa system, considerable barriers remain". Foreign scientists must still submit to lengthy security review, the group said. And the government is considering a new set of rules that could limit researchers' activities in the lab (see *Nature* 435, 4; 2005). The group is asking that those rules are not brought in.



Hazy prospects: the future of agreements on greenhouse-gas emissions is clouded in doubt.



AVIAN FLU: Are we ready?

P. BRONSTEIN/GETTY IMAGES

Trouble is brewing in the East. A highly pathogenic strain of avian influenza is endemic in southeast Asia. Many millions of chickens have been culled, but there is a persistent reservoir in domesticated ducks and wild birds. The H5N1 virus isn't going to go away. And each time it emerges, people can be infected.

H5N1 first reared its head in Hong Kong and southern China in 1997, killing six. Since late 2003, it has led to the deaths of more than 50 people in Vietnam, Thailand and Cambodia.

The stage is set for the emergence of a fresh human influenza pandemic. These occur when a virus to which most people have no immunity, usually an avian strain, acquires the ability to transmit readily from person to person. H5N1 hasn't yet gained that ability — and hopefully, it will not.

But if it does, the virus could spread across the globe within months. The consequences are difficult to predict. We're unlikely to be as lucky as in 1968, when the relatively mild H3N2 virus killed some 750,000 people worldwide. But the real nightmare scenario is a re-run of the H1N1 flu pandemic of 1918, which left as many as

40 million dead. Standards of health care have improved a lot since then, which will help. But if a pandemic strain were to retain H5N1's current extreme pathogenicity, a similar toll can't be ruled out.

This week, *Nature* devotes its News Feature and Commentary pages to a detailed consideration of the risks posed by avian flu, and how well we are prepared to deal with it. In the pages that follow, our reporters examine nations' capacity to produce a vaccine against a pandemic strain, and the adequacy of global stockpiles of antiviral drugs. They do not paint an encouraging picture.

Repeated warnings about the international community's failure to respond to the pandemic threat have fallen on deaf ears. So in our opening News Feature, we use the benefit of fictional hindsight to throw the issues into starker relief, describing a future pandemic through the weblog of a journalist in the thick of things. This is fiction, but not fantasy — the storyline was drawn up in consultation with those who could soon be dealing with the situation for real.

In our extended Commentary section, starting on page 415, experts who are

grappling with the issues tackle some hard questions. Which nations are ready, and which are not? David Ho asks if China is in a better position to cope with new microbial threats since the 2003 SARS outbreak. And Anthony Fauci outlines what US researchers are doing to develop vaccines and drugs.

Asian countries are the most immediately vulnerable. Robert Webster and Diane Hulse address the possibility of controlling flu outbreaks in these nations at source, pointing to two examples of successful intervention to wipe out the disease in poultry — in Hong Kong in 1997 and more recently in Thailand. Joining up the dots between animal and human health is also the concern of Albert Osterhaus and his colleagues. They propose a permanent global flu task force to strengthen coordination among agencies on the ground.

If we are fortunate, we may still have the time to take these messages on board. As Michael Osterholm warns in his Commentary, a flu pandemic could bring human tragedy and a global economic catastrophe. Let's hope world leaders heed the warnings.

Peter Aldhous, chief news & features editor.
Sarah Tomlin, commentary editor.

The flu pandemic: were we ready?

HOME 26.12.05 28.12.05 29.12.05 30.12.05 31.12.05 25.01.06 02.02.06 18.02.06 20.02.06 27.02.06 17.05.06 LINKS

Welcome to my weblog. I'm Sally O'Reilly, a freelance journalist based in Washington DC. I've been researching a book on pandemic preparedness. But now the time for preparation has run out.



• 26 DECEMBER 2005 IT'S AN EMERGENCY — OFFICIAL

President George Bush has just addressed the press in the East Room of the White House. Here's the transcript: "At this hour, the [World Health Organization](#) has declared a full-scale pandemic influenza alert, with person-to-person spread lasting more than two weeks in Cambodia and Vietnam. During previous influenza pandemics in the United States, large numbers of people were ill, sought medical care, were hospitalized and died. On my orders, the [Department of Homeland Security](#) and the [Department of Health and Human Services](#) have today implemented the nation's draft [Pandemic Influenza Response and Preparedness Plan](#). It will serve as our road map, on how we as a nation, and as a member of the global health community, respond to the pandemic. We are ready. Thank you, and may God bless America."

Ready, my ass! I've reported on avian flu for almost a decade. The first thing I did on hearing Bush's address was to get on my cellphone to my husband, Jonathan. I told him to pack some bags and get ready to take the kids to my mother's house in Florida. "Remember all that stuff I told you about how a bird flu pandemic might hit the United States? Well, I think it's about to happen."

• 28 DECEMBER 2005 JOURNEY TO THE SOURCE

Hanoi, Vietnam. I'm exhausted, and I can still taste the disinfectant they sprayed inside the [Doctors Without Borders](#) plane. I'm at the Bach Mai Hospital. It was here, three weeks ago, that what they're calling the 'Hanoi index case' fell sick. A Malaysian on business, he was transferred to a hospital in Hong Kong, where he died. Samples sent to labs in the WHO Global Influenza Surveillance Network showed he was infected with an H5N1 avian flu virus, but one that differed from earlier isolates. It had mutated.

But he won't have been the first patient with this mutated strain. As early as October there were hundreds of human H5N1 cases in the countryside south of here, but only a handful got picked up. Most went unnoticed by health authorities. Surveillance for human cases of flu in Vietnam has been patchy, and DNA diagnostic tests unreliable. WHO calls for more international funding were ignored. Now the virus has had three months to spread, pick up mutations and get better and better at jumping between humans.

What's weirdest is that there weren't any declared outbreaks of bird flu in chickens here recently. Farmers weren't exactly queuing up to declare cases, though. There'd been talk of setting up a global fund to help them cope with eradicating avian flu, to compensate them for lost trade. But it came to nothing. Then again, perhaps the virus came from ducks, which can be infected without showing symptoms.

• 29 DECEMBER 2005 LIFE BEHIND THE MASK

Today, I hooked up with the 15-person international team from the WHO's [Global Outbreak Alert and Response Network](#). They're like the cast in that movie *Outbreak*, about some monkey virus.

We've got epidemiologists from the CDC — the US [Centers for Disease Control and Prevention](#) — along with mathematical modellers from Imperial and Emory, and virus hunters from the Pasteur Institute in Paris and the Robert Koch Institute in Germany. They're here to help hospitals control infection, and strengthen surveillance for human cases. Another team is doing the same in Cambodia. Across the world, health authorities are ramping up surveillance, trying to spot and isolate any exported cases as quickly as possible. They've grounded all commercial flights to and from the region. The chaos is way worse than with SARS.

Second evening here. The N95 face masks, which the WHO has advised us to wear, are the worst part. Your glasses steam up and you feel half-suffocated. I only take mine off to eat and drink. The team has a web video conference via a high-bandwidth satellite connection with WHO headquarters in Geneva. Its Department of Communicable Disease Surveillance and Response is coordinating the international response. Poor guys, there's just a handful of them.

They run through the latest stats. Here we go: 1,800 cases in Cambodia, 1,100 in Vietnam. Uh, oh ... six suspect cases in Tokyo and Johannesburg. So much for the flight bans. Overall, the mortality rate is 9%. That's nasty — worse than 1918. But it'll come down, as there's probably loads of asymptomatic cases.

The labs have finished sequencing the virus and we now have a template for an H5N1 vaccine. But it won't be ready for months. So for now, the WHO is trying a long shot, known as targeted antiviral prophylaxis.

http://www.nature.com/news/2005/050523/full/435400a.html

Basically, the idea is to blanket bomb all index cases, their households and people in the immediate vicinities with antiviral drugs such as Tamiflu. Computer models predict that if we do this, we might just stop the pandemic in its tracks. But there hasn't been enough modelling, and now we're doing the experiment for real.

Continued modelling will be vital, though, to work out how to deploy the limited supplies of Tamiflu we've got, and how long we need to treat people for the drug to work. Geneva informs us that the WHO international stockpile contains just 120,000 pills. WHO officials have been on the phone today with countries that have national stockpiles.

The politicians know that stopping the pandemic at source would be the best solution. But they're reluctant to donate drugs, as they'll have less for their own citizens if this approach fails. No point asking the United States — they've only got enough pills for 1% of the population. Britain and France have enough for a quarter of their populations. Will they spare us any? Will they hell.

• 30 DECEMBER 2005 GETTING TO KNOW THE ENEMY

Geneva announces that the latest epidemiological studies say that the virus seems to have a 'basic reproductive number', or R_0 , of between 1.4 and 2.0. This means that one case on average infects only one or two people. So if we can detect cases quickly and treat them and their contacts, the models suggest we could contain the virus most of the time. At the least, that might slow the pandemic and corral it in that region for a few months. That would win time to get a vaccine.

But we know there is a very short window. As time goes by, this virus will get better and better at transmitting between humans, and the R_0 will increase. If it goes above 3, there's no way we'll contain it.

The latest news from Cambodia cheers us up. There's a slowdown in new cases. Control efforts seem to be keeping the lid on the virus there. But here it's a different story: the team is having difficulties finding and isolating contacts of patients in this crowded city.

This flu moves much faster than SARS because its incubation period is just two days. People are spreading the virus the day before they get sick, and asymptomatic patients without even being visibly ill. Tamiflu needs to be administered within two days of anyone showing symptoms.

As I wandered through the streets this afternoon, it wasn't looking good. People are walking around Hanoi coughing and spluttering. They've closed the schools, which is the right thing to do, but what are all the kids doing? Hanging out downtown enjoying the unexpected holiday.

• 31 DECEMBER 2005 SIX MONTHS TO A VACCINE!

Vaccine teleconference. There are 125 people — companies, regulators, scientists — hooked in, each with their own agenda. It's impossible. There's a lot of talk on whether the six-month delay before there is any vaccine can be shortened. Scientists had been working on methods of growing virus for the vaccine in large vats of cultured animal cells instead of eggs. That could cut the delay to maybe three months. But progress had been held up by US [Food and Drug Administration](#) concerns over the safety of the cell lines. In any case, it would probably take at least two years before the existing factories could be switched over.

So we're stuck with eggs. A fast-track FDA approval for an H5N1 vaccine is under way. Fortunately, the US Department of Health and Human Services last year funded Sanofi-Pasteur to test a 'mock' H5N1 vaccine, using antigens from an earlier strain. So we don't need to start the approval procedure from scratch for the pandemic strain. We've gained some time.

But US production capacity — one factory — is only enough to cover up to 90 million people. The situation is better in the European Union: it can probably produce enough to cover 30% of its 450 million people. The predictable news is that every vaccine-producing nation has just nationalized its supply to serve its own citizens first. The 'have-not' countries aren't going to get any vaccine.

There's a lot of hindsight and recrimination at this meeting. The United States only tested vaccines at standard doses. Testing a vaccine containing an immune-boosting adjuvant might have allowed it to be diluted eightfold. Even with existing world production capacity, that would have let us produce 7.2 billion shots, enough to treat half the world's population. Now it's too late.

• 25 JANUARY 2006 ESCAPING FROM HELL

Apologies for the long delay in posting. The past few weeks have been chaos. I was out with WHO teams from dawn to dusk as they tried in vain to stamp out the outbreak with drugs. People fell sick all over Hanoi and 1 in 50 of them died. Many of the worst affected felt fine in the morning, but were dead by lunchtime — blue in the face, gasping for air. At the overcrowded hospital, I saw victims collapsing, suffocating in their own lung fluid, blood streaming from their noses and gums. Others had longer ordeals, tortured by encephalitis as the virus ate into their brains, or overwhelmed by multiple organ failure. Panicky authorities transported corpses out to the fields by truck and burnt them on open pyres.

In a desperate last attempt to quell the outbreak, the WHO took what drugs it had left here and blindly treated whole sections of the city where transmission was most severe. The army was supposed to enforce quarantine, but many of them were sick as well, or had joined the exodus from the city. The fleeing people inevitably spread the disease to the countryside.

In a few days' time, the Vietnamese are supposed to celebrate Tet, the lunar new year festival. It's traditional to eat chicken — but not this year. My plane leaves tonight. I feel like I'm escaping from hell.



P. BRONSTEIN/GETTY IMAGES



• 2 FEBRUARY 2006 THE VIRUS SPREADS

Today, I was at a press conference at the [National Institutes of Health](#) in Bethesda. A guy from the CDC pointed to a giant screen, a map of the world dotted with red pixels. He said that they'd reckoned the virus might hit in two or more waves up to eight months apart, as in past epidemics. They'd hoped the first pandemic strain of H5N1 might be poorly contagious, and come back again with a vengeance after it had picked up more infectivity. By that time we might have had a vaccine. That was just a hunch, though. And it was wrong.

The mild pandemic in 1968 took almost a year to cross the globe. This one probably started around October. So we're now almost four months in. Look at that map! With the huge increase in passengers travelling by air, it's already lodged in 38 cities around the globe. The outline of Asia is barely visible beneath the swarm of red pixels.

Now the virus is in coastal cities on both sides of South America. It hit Europe two weeks ago, ripping through Paris in just 11 days. In the French capital alone, there were 2.5 million cases and 50,000 dead. That's par for the course — infection rate 25% and mortality 2%, similar to the 1918 pandemic. Extrapolate these numbers, and we're going to have over 30 million dead worldwide. In poor and densely populated countries like India, it could be worse.

Where's next, I asked. Based on passenger data — which had to be prised from the airlines — one epidemiologist was willing to make a guess. "Within two weeks, there." He traced his finger from San Diego to Los Angeles, up to San Francisco. Within another three to four weeks, it'll be the turn of the conurbations along the eastern seaboard.

• 18 FEBRUARY 2006 THIS CAN'T BE HAPPENING

The United States is battened down before the storm. The government has outlawed all gatherings in public places. In past pandemics that never worked. But epidemiologists say that if we do it early on, it might slow the spread. Modelling also suggests that closing schools and universities is especially important as teenagers and young adults are among the worst hit. We just need to stop them from hanging out elsewhere. Stay at home, is the message blaring from every TV screen.

On [CNN](#) it's now round-the-clock coverage, with a red 'Pandemic' banner running across the bottom of the screen. "We're in the twenty-first century, and they're telling us about how to wash our hands properly, and practise 'respiratory etiquette'," exclaims Jonathan. "Why aren't there drugs? And I can't believe there's no vaccine. This can't be happening in America."

• 20 FEBRUARY 2006 AMERICA SHUTS DOWN

The [Commissioned Corps of the US Public Health Service](#), the nation's uniformed force of health professionals, has just been mobilized. The [US Northern Command](#) is in charge of the military response. Soldiers are setting up triage centres, anticipating overflowing emergency rooms and morgues. Images are coming in of tent cities being erected in New York's Central Park. Wards are being installed in schools and churches. Troops are on the streets. "There's going to be civil unrest," a general informed me on the phone this morning.

The CDC is in charge of national influenza surveillance, but it's a nightmare now. This is the peak season for ordinary flu, sparking false alarms and panic. Scant supplies of Tamiflu are being reserved for medical first responders, and essential services. (Stocking cash machines is an essential service, we learn.)

There's a lot of looting going on in pharmacies, but to no avail. The drugs are being distributed in convoys, with military jeeps in front and behind. Masks costing a dollar are being sold on street corners for \$20. E-mailed ads for counterfeit drugs are filling up my inbox.



• 27 FEBRUARY 2006 EVERYONE FOR THEMSELVES

I watch the scenes of a society descending into chaos from the relative security of my mother's isolated home. Red tail lights snake to the horizon as people pour out of the cities. Half the doctors haven't turned up for work; many are either ill, or caring for loved ones.

Who should get the few mechanical respirators that can mean the difference between life and death? The youngest, or those with the best chances of pulling through? "Our leadership must be prepared to make calculated decisions that will force raw prioritization of life-saving resources," explains a colonel on CNN.

• 17 MAY 2006 THE DUST SETTLES

The pandemic was [declared over](#) today. H5N1 will be back next year, or before that, as it replaces the existing seasonal flu strains. But by then, those who have recovered from this bout will have immunity, and we will have a vaccine. Pandemics move faster than governments or international bureaucracies, and the cost is hundreds of billions of dollars more than it would have been had we tackled avian flu in Asia in the first place, and invested in flu research. For millions of families, the cost isn't measured in dollars.

Watching all that military hardware on the streets made me think. We imagined we could encourage pharmaceutical companies to develop innovative vaccines and drugs by offering 'incentives' or modest subsidies. When the military knows it needs a fighter aircraft, it doesn't offer incentives to Lockheed Martin or Boeing. It pays them through procurement to develop the weapon to the specifications it wants.

Were we ready? Ready, my ass!

Sally O'Reilly's blog was written by Declan Butler, *Nature's* senior reporter in Paris.

J.S. APPLEWHITE/AP

Is this our best shot?

We have the means to make a vaccine against pandemic flu. But quarrels over money, science and politics mean it could come too late, says **Erika Check.**

On a Tuesday afternoon in early April, a young woman with intense green eyes sits in a hospital examining room in Rochester, New York. Pressing a ball of cotton into the crook of her elbow, she explains why she has just allowed a nurse to draw two and a half teaspoons of blood from her slender arm, even though she hates needles. Oksana, 29, is testing a vaccine that could save humanity from bird flu. "Each of us has to do something to stop it — as much as we can," she says.

Oksana — who asked *Nature* not to use her real name — has volunteered for a trial to test a vaccine against the H5N1 avian influenza virus. The study, funded by the US National Institutes of Health (NIH), includes 450 volunteers like Oksana. But their efforts could go to waste unless health officials, world leaders, scientists and businesses find fast answers to a mass of difficult issues. Money troubles, politics and hiccups in production processes could stymie the development of a vaccine to protect us from a flu pandemic.

Vaccination against the common varieties of influenza that have been infecting people for years is nothing new. In 2003, drug companies sold some 292 million doses of the seasonal flu shot¹. But a pandemic strain that has crossed over from birds will be so different from com-

mon flu that the immune response created by our run-of-the-mill vaccines will be useless.

A dangerous avian flu virus such as the H5N1 strain could morph into a pandemic virus in two ways. The virus could mutate so that it can be passed between people, or it could exchange genes with a common human flu strain (see 'Deadly combinations', overleaf). Once a pandemic strain is born, researchers will find themselves in a frantic race to create a vaccine that is effective against it.

But that could take months — and in the meantime, a pandemic virus could circle the globe. So scientists are now gearing up to test human vaccines against H5N1 and H9N2, the two most threatening strains of avian flu.

If a pandemic strain evolves from either of these viruses, it may not be an exact match for the vaccines we are testing now. But hopefully it will look similar enough for the vaccine to provide some protection.

At least ten such trials are scheduled this year across Australia, Canada, France, Germany and Japan. Thailand and Vietnam are drawing up plans for trials, and the United States has already begun its studies with Oksana and her fellow volunteers.

This is welcome news. Just seven months ago, Klaus Stöhr, the chief influenza expert at the World Health Organization (WHO),

warned that the world was asleep at the wheel. Back then, only two countries had plans to test a human vaccine against avian flu, and the WHO called a meeting of vaccine makers and government officials to sound the alarm. "There is currently too little momentum in the development of pandemic influenza vaccines," Stöhr said at the meeting, held in Geneva. "We had three pandemics in the last century, and there is no reason to believe there won't be one in this century."

A shot in the arm

Results from earlier, small trials of H9N2 vaccines suggest that the human immune system might respond well to an avian flu vaccine only if it gets an extra kick from an ingredient called an adjuvant²⁻⁴. Adjuvants are chemical additives that seem to 'irritate' the immune system, dramatically boosting the response to a vaccine. But most countries have not approved the use of adjuvants in flu vaccines, so they need to undergo additional testing — and that will take extra time. The US government has said that it will pay for tests of an H5N1 vaccine boosted by an adjuvant. But for now, the vaccine trial in which Oksana is enrolled — which doesn't include an adjuvant — is the only game in town.

To make a flu shot, scientists usually inject flu viruses into fertilized chicken eggs, let the viruses copy themselves, and then kill them with chemicals. But before this step, researchers have to modify the viruses so



"We had three pandemics in the last century — and there is no reason to believe there won't be one in this century." — Klaus Stöhr

A. WONG/GETTY IMAGES

that they grow well in eggs. Traditionally, researchers have done this by infecting eggs with two different flu strains: a 'wild' virus, which causes illness, and a lab virus, which grows well in eggs. If everything goes according to plan, the two viruses mix their genes and reassemble into a vaccine strain that works against the wild flu virus. But this takes time and luck — and if your luck is out, vaccine production can be delayed by months.

Today, it is possible to save time by using a technique called reverse genetics, perfected for flu viruses just five years ago⁵. This involves stitching flu-virus genes into loops of DNA called plasmids. The plasmids can assemble into whole flu viruses in the lab. By using reverse genetics, scientists can create exactly the types of virus they need, avoiding the lottery of natural reassortment.

Researchers have already used reverse genetics to create candidate vaccines against H5N1 flu^{6,7}. Each flu virus is named after the types of two proteins that make up its outer coat (see Graphic, below) — haemagglutinin (H) and neuraminidase (N). The genes for these proteins are constantly mutating and come in many different varieties. To make their H5N1 vaccines, the researchers first altered the haemagglutinin gene from an H5N1 strain to make it less deadly, and then applied reverse genetics to create their mix-and-match vaccine strains.

Oksana's trial is testing one of these vaccines, developed by a team at the St Jude Children's Research Hospital in Memphis, Tennessee. At the NIH, meanwhile, researchers led by Kanta Subbarao and Brian Murphy are creating a library of vaccines against a range of different avian flu strains, including H5N1 and H9N2. Unlike conventional flu vaccines, these would contain live, weakened viruses in a nasal spray, instead of dead viruses in a shot. The researchers hope that this will generate a stronger immune response.

Back to square one

But if a pandemic strain is vastly different from the vaccine strains that have already been tested, scientists will have to make a new vaccine from scratch. The St Jude researchers say that they could make a vaccine strain just four weeks after they get their hands on a sample of the pandemic strain⁸. But there's a problem: reverse genetics has been patented, so companies that make the vaccine would have to pay royalties to the patent holders. Companies are reluctant to do this, but scientists working in the field say that industry is trying to hammer out this issue now.

Even if the intellectual-property issues are resolved, it will be very difficult to step up global vaccine production to make enough to halt a pandemic. Vaccine companies currently make 300 million flu shots a year. But in a pandemic, we could need billions of doses. So why haven't vaccine makers leapt to fill this gap?

The answer is money. Flu vaccines are simply not a lucrative prospect for drug companies, which can make much higher profits on blockbuster drugs. And vaccines are risky: anything injected into a healthy person can end up doing more harm than good, leading to costly lawsuits and bad press. What's more, a flu pandemic might never hit, so business leaders are reluctant to spend money on new factories that might never be used.

So at present, vaccine manufacturers have limited production capacity, which could be further constrained by the supply of lab-standard fertilized chicken eggs. Growing the virus in huge metal fermenters containing a soup of cultured cells could be faster than using eggs. But retooling entire factories is an expensive business. "We are talking about a totally unpredictable, very rare event, so it's difficult to commit a company to these

"After the pandemic occurs, many scientists will be held accountable for what we did or didn't do to prevent it." — Michael Osterholm

INFLUENZA'S DIRTY TRICKS

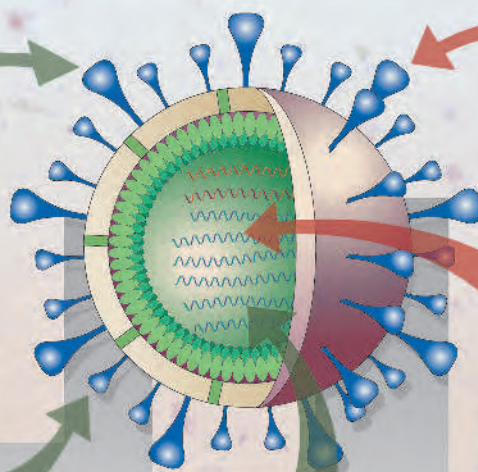
FIGHTING BACK

Killer key

The flu virus uses the protein haemagglutinin (H) to enter the host's cells. This protein constantly mutates to stay one step ahead of the human immune system.

Spreading infection

The protein neuraminidase (N) cuts newly formed flu virus free from the host cells to spread through the body. It also mutates rapidly.



Pandemic engine

Influenza's genome is split into eight different pieces, each carrying a single gene. If two different strains infect the same animal, they can mix genetic material and produce a radically different strain that can trigger a pandemic.

Vaccine

Antibodies against haemagglutinin and neuraminidase are essential to protect against infection and spread, so any vaccine must contain these proteins.

Designer viruses

Scientists can exploit influenza's piecemeal genome to engineer viruses ideal for use as vaccines. They can isolate the H and N genes from a dangerous pandemic virus and alter them to make them safer. They then mix these with the genes of another flu virus that is easy to grow in the lab. These genes assemble into new viruses that can be used in vaccines.

Deadly combinations

Dog kidney cells can help experts to assess whether a deadly avian flu virus will mix its genes with a common human strain to create a virus that could kill millions. In these cells, virologists at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, are now running tests using various genetic combinations of the H5N1 avian flu virus and a common human flu strain.

In January, under tight security, CDC virologists began to mimic natural genetic 'reassortment' using reverse genetics (see main text). Using the H5N1 strain as a backbone, they have been substituting various combinations of eight genes from the H3N2 human flu virus.

There are 254 possible combinations, so early experiments are simply screening them to see whether they can survive in mammals — which is where the canine cells come in. "We're trying to approach this in a systematic way," explains

Nancy Cox, who heads the CDC's influenza branch.

In the autumn, Cox's team plans to test the most viable and dangerous blends in live animals, to see which ones are readily transmissible. The idea is to get a preview of what a pandemic strain might look like, so workers in the field can be primed to look for novel viruses that may pose a particular threat.

Virologist Albert Osterhaus of Erasmus University in Rotterdam, the Netherlands, expects to get approval from national authorities in the next two months to begin similar research. The animal studies will prove the most challenging, he says: "Transmission experiments are notoriously difficult. You have to mimic the normal human situation."

Osterhaus also aims to conduct transmissibility studies of other strains of bird flu, including H7N7 — which jumped to people in the Netherlands in 2003, causing symptoms including conjunctivitis, and killing one person. "The



CDC

Nancy Cox is seeking the likely form of a pandemic strain of flu.

attention the H5 virus has received is logical, but we should not forget that other avian viruses could do the same thing," he says.

Such research is intensely controversial, because of the potential for giving bioterrorists a recipe to create mayhem. Cox and Osterhaus point out that the genetic engineering involved is beyond the reach of most terrorist

groups. "But if we found something that was truly horrific, we might not release that information," Cox adds.

The bigger danger, in fact, is that nature will do the job first. Cox and Osterhaus know they're in a race against time that they can't be sure of winning. "We just don't know when a pandemic strain will strike," says Osterhaus.

Roxanne Khamis

preparational approaches," says Norbert Hehme, who manages GlaxoSmithKline's vaccine facility in Dresden, Germany.

Governments could help to coax drug companies back into the flu-vaccine business by

increasing the use of shots against common, non-pandemic flu. The Canadian province of Ontario, for example, gives standard flu shots to its entire population each year. And last August, the US government added flu shots to the list of vaccines recommended for all infants. These sorts of steps are reassuring companies that someone will buy their products. Bruce Gellin, head of the National Vaccine Program Office at the US Department of Health and Human Services, says that more companies are discussing the possibility of making pandemic vaccines. "The fact that manufacturers are interested says that it's a much more attractive marketplace now," he says.

But as companies follow the money, some countries will be left behind. In 2003, nine rich nations, led by Japan and the United States, used 62% of the world's influenza vaccines¹. But bird flu is most rampant in poorer countries that do not buy a lot of flu shots. Although Vietnam and Thailand are planning clinical trials of influenza vaccines, the plans are preliminary. So there is a high risk that the countries at ground zero will be defenceless in the early days of a pandemic.

That would be a tragedy for the countries concerned — and could threaten us all. If rich nations lock down their vaccine supplies and watch pandemic flu rage through southeast Asia, they will hasten its spread across the globe. Public-health experts say that governments should pledge now to share their vac-

cine supplies. But no formal talks have been scheduled. "Some ideas exist on how to organize sharing, but the reality is that no country with a vaccine producer has come forward with a proposal," says Stöhr.

Experts say that we must tackle these problems soon — before it's too late. In Washington DC last month, Michael Osterholm of the University of Minnesota spelled out the issues to a room full of experts gathered by the US Institute of Medicine. "After the pandemic occurs, there will be a post-9/11-like commission," Osterholm said, referring to the high-level US panel that investigated preparedness against the 2001 terrorist attacks. "And many scientists will be held accountable to that commission for what we did or didn't do to prevent a pandemic." The same will be true for government officials and business leaders who control our ability to deploy an effective pandemic vaccine.

Erika Check is Nature's Washington biomedical correspondent.

1. Fedson, D. S. *J. Pub. Health Pol.* **26**, 4-29 (2005).
2. Hehme, N., Engelmann, H., Kunzel, W., Neumeier, E. & Sanger, R. *Med. Microbiol. Immunol.* **191**, 203-208 (2002).
3. Stephenson, I. et al. *Lancet* **362**, 1959-1966 (2003).
4. Wood, J. M. & Robertson, J. S. *Nature Rev. Microbiol.* **2**, 842-847 (2004).
5. Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G. & Webster, R. G. *Proc. Natl Acad. Sci. USA* **97**, 6108-6113 (2000).
6. Subbarao, K. et al. *Virology* **305**, 192-200 (2003).
7. Lipatov, A. S., Webby, R. J., Govorkova, E. A., Krauss, S. & Webster, R. G. *J. Infect. Dis.* **191**, 1216-1220 (2005).
8. Webby, R. J. et al. *Lancet* **363**, 1099-1103 (2004).



Out of luck: creating a flu vaccine by growing viruses in eggs has a relatively high failure rate.

What's in the medicine cabinet?

Drugs that could lessen the death toll in a flu pandemic do exist. But global stockpiles are too small, and the countries at most immediate risk are among the worst prepared. **Alison Abbott** reports.



The pharmaceutical company Roche didn't have huge commercial expectations for its influenza drug oseltamivir when it was licensed under the brand name Tamiflu in 1999. Flu is a fact of life, and doctors have been advising aspirin, hot lemon and bed-rest for generations. In most countries they continue to do so, reserving the drug for vulnerable groups such as the elderly.

But in the past year or so, Roche has quadrupled its Tamiflu production capacity. The reason: developed countries are now stockpiling the drug against the threat of a pandemic flu virus that could arise at any time. Given the difficulty of rapidly producing an effective vaccine (see page 404), drugs will be the first line of defence. But even after Roche's moves to boost Tamiflu production, experts say that global stockpiles are woefully inadequate.

What's more, no one knows for sure the answers to several key questions. How many deaths could antiviral drugs prevent? To what extent would they slow the spread of a pandemic? Could they, as some mathematical modellers claim, even stamp out the disease as it emerges? "There is a lot of uncertainty, but that is no reason not to plan their use," says Marc Lipsitch, an infectious-disease epidemiologist at the Harvard School of Public Health in Boston.

Home guard

Although the pandemic will be global, defence plans are so far strictly national. Thanks mostly to prodding by the World Health Organization (WHO), about 50 countries have drawn up pandemic-preparedness plans. Most are still very sketchy, but include strategies for stockpiling antiviral drugs. Only a handful of nations, including Britain and Canada — but notably not the United States — have given their plans legal status.

Worryingly, the list of relatively well-prepared nations includes few of those countries in Asia where a pandemic strain is most likely to emerge. Historically, the WHO has found it hard to persuade even rich countries to produce a pandemic plan: many governments

Rows of antiviral flu drugs fill Roche's warehouse — but stocks would not meet demand in a pandemic.

ROCHE

have proved reluctant to pay for a stock of drugs that may not be used during their terms of office. Only now that the alarm bells are ringing about the H5N1 avian flu virus have official minds been focused.

Experts agree that Tamiflu is the best of the four currently available anti-influenza drugs. A course costs between US\$10 and \$30, but national stockpilers have negotiated prices in the lower range. Roche is also making the powdered active ingredient available at a cheaper price than tablets. The powder would be dissolved in water and drunk when needed

— nasty-tasting but still effective, and stable in solution for several days.

Tamiflu, and the chemically related zanamivir, marketed by GlaxoSmithKline as Relenza, belong to a class of drugs called neuraminidase inhibitors. They do not eliminate the virus, but they reduce its release from infected cells by blocking a key viral enzyme. If taken within 48 hours of the onset of symptoms — the earlier, the better — they reduce the duration of symptoms by at least a day^{1,2}. They also limit the severity of symptoms in non-pandemic flu: patients succumb less

frequently to acute bronchitis or pneumonia¹. That should be good news if the same applies to a pandemic strain, as patients who cough less will spread the virus less effectively.

Relenza is less helpful because it has to be taken by inhaler, which is not very practical if a patient's breathing is impaired. But both neuraminidase inhibitors have so far generated few problems with drug resistance: mutations in the flu virus that confer resistance seem rare, and generally seem to weaken it³. (But an H5N1 virus sample from one Vietnamese patient has recently been shown to be less susceptible to Tamiflu, so experts are not complacent.) Side effects are also mild, and the drugs can be kept on the shelf for at least ten years without losing their activity.

The older, off-patent drugs amantadine and rimantadine belong to a different class and interfere with a viral protein called M2, which stops the virus from entering its target cells. They seem to be as clinically effective as the neuraminidase inhibitors, but resistance arises very rapidly and the drugs can have disturbing side effects, including psychotic episodes. Although such reactions are rare, they would be highly unwelcome in the already panicky atmosphere of a flu pandemic.

Fair treatment?

Indeed, the potential for social unrest is a major concern for those laying pandemic plans. And demand for Tamiflu could exacerbate the problem. Who will, and who will not, be treated with this scarce but valuable resource? "It's not easy — we know there won't be enough for everyone," says Theresa Tam of the Public Health Agency of Canada. Britain, which is among the best-prepared countries, has ordered enough for about 25% of its population; Canada has stocks for just over 5% of its people; the United States currently cannot even cover 1%.

In practice, a significant proportion of supplies might be used for prophylaxis of health-care workers — for up to two months as the influenza wave passes through — leaving less for treating the sick. "It is not a happy situation," says Klaus Stöhr, the WHO's chief influenza expert. Canada, wary of the potential for a public backlash if health workers were perceived to be saving their own skins, included an ethicist on its Pandemic Influenza Committee.

The WHO recommends that antiviral drugs should be available for the early treatment and prophylaxis of "those groups at highest risk of infection" and "essential workers". But defining these people, and matching their number to the doses available, is difficult.

Ultimately, how you define your strategy depends on what you want to achieve, says clinical virologist Fred Hayden of the University of Virginia in Charlottesville. Most countries are aiming to keep the death toll as low as possible, but for others, maintaining the economy may be at least as high a priority. So the definition of essential workers will vary. Those deemed non-essential will be able to do little



D. MACKINNON/GETTY IMAGES

Masking our ignorance

When SARS, or severe acute respiratory syndrome, hit the cities of Asia in 2003, one product was in hot demand: the N95 face mask.

These cup-shaped masks fit snugly on the face and filter out particles smaller than a few hundred nanometres across. Flu viruses are smaller than this, but are often coughed or sneezed out in larger droplets. Official advice on whether N95 masks offer protection against flu is confusing, to say the least.

The World Health Organization (WHO) recommends that people at the highest risk — health-care workers and the families of those infected with the disease — wear N95 masks, which retail for about US\$1 each. Many national health agencies are following this advice. Given that flu is largely

transmitted in droplets, N95 masks should be of some value, suggests Klaus Stöhr, the WHO's chief flu expert. But the US Department of Health and Human Services takes the opposite view. Its pandemic plan, released in August 2004, states: "N95 respirators, which would be recommended for infections with airborne spread such as tuberculosis, are not required for influenza."

Experience with the SARS virus, which is roughly the same size as flu viruses, and seems to be spread in a similar manner, shows that N95 masks aren't completely reliable. Researchers in Canada reported that nine health-care workers developed SARS from a single patient despite using the masks and other recommended infection-

control procedures⁵.

And some experts worry that the emphasis on N95 masks, rather than simple polypropylene surgical masks that cost a few cents each, is misguided. Wing Hong Seto of the Queen Mary Hospital in Hong Kong led a study on health-care workers during the SARS outbreak, and showed that cheaper surgical masks were effective in helping to prevent transmission as part of a suite of infection-control measures⁶.

Seto warns that all masks can pose a threat if reused or not disposed of carefully. "They could be contaminated with droplets that are then passed on to the hands," he says. "They should be used only once." In poorer countries, cheap surgical masks may be a more viable disposable option.

David Cyranoski



VIPs: in some countries, will drugs be earmarked for 'essential' construction workers?

but don a protective face mask — which provides no guarantee of safety (see 'Masking our ignorance', opposite).

But the biggest challenge to any plan is the intrinsic biological uncertainty: just how nasty will a pandemic virus be? "We have so many unknowns — about how many people of what age groups would get ill, just how ill they would get, just how fast the virus would transmit — so it is hard to be firm about the best strategy for prioritizing treatment groups," Hayden says.

Of course, the larger the stockpiles, the easier the choices will be. This is why the WHO is using all its persuasive powers to get governments to place orders now. Once a pandemic breaks out, it will be too late. Roche has promised not to profiteer by hiking prices during a pandemic, but it is not simply a question of money. The firm has no spare production capacity and batches take up to a year to make.

In addition to encouraging stockpiling,

experts are trying to find other ways of driving up the supply of antiviral drugs. They argue strongly, for example, for the wider prescription of antivirals against non-pandemic influenza. "This would allow companies to increase their routine manufacturing capacity without fear of losing money," says Stöhr. Other countries should follow the example of Japan, which consumes three-quarters of the Tamiflu prescribed each year, he argues. Most of the rest is used in the United States, with only 3% being prescribed in the rest of the world.

"It would be very good for physicians in these other countries to have experience with the drug before a pandemic arrives, so that they learn how best to treat patients," agrees Hayden. It's important for patients to be hit with the drug early, he says, but doctors may accept this only through clinical experience. The wider use of antivirals during annual flu epidemics would also stimulate companies to develop new drugs. This is currently not a priority for the pharmaceutical industry because the market is too small.

Unknown quantity

There is still of plenty of work to be done to further our understanding of Tamiflu's pharmacology. "There are gaps in our knowledge that we need to fill so that physicians can use it more effectively in a pandemic," says Hayden. For example, Tamiflu is not licensed for infants under one year old, because of the ethical difficulties of running trials in very young children — yet this age group proved exceptionally vulnerable in the severe pandemic of 1918.

Pharmacologists also want more biological data on patients who are treated with Tamiflu after being infected with the H5N1 virus now circulating in Asia. This will help them optimize dosing regimes. They complain that not enough is being done to gather these data from the relatively few patients who have so far been given the drug. Animal studies would also help, but this has similarly not yet been made an official priority.

"There are gaps in our knowledge that we need to fill so that physicians can use Tamiflu more effectively in a pandemic." — Fred Hayden

Animal models could be used to investigate the use of Tamiflu in drug combinations, which may help avoid any problems with resistance. Pandemic planners are considering stockpiling amantadine and rimantadine as back-ups, despite their disadvantages, because they are cheap and were shown to have some prophylactic activity in the milder 1968 pandemic⁴. Now is the time to begin investigating the merits of using both major classes of anti-flu drugs together, says Hayden.

Finally, experts urge that the few new candidate drugs coming up should be given serious consideration, even if they don't seem ideal. For example, peramivir, another neuraminidase inhibitor was developed by BioCryst Pharmaceuticals of Birmingham, Alabama, but abandoned because it has to be injected. Nevertheless, its very long half-life in the body means that it needs to be given only once or twice a week and so might be useful prophylactically.

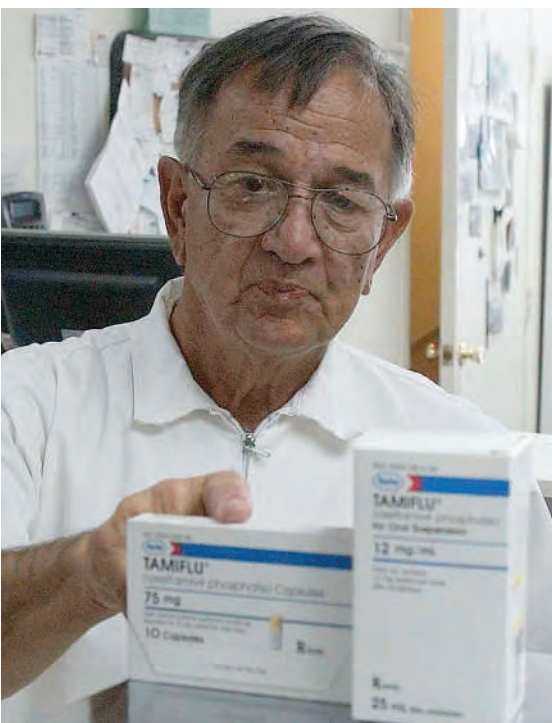
For most developing countries, meanwhile, creating a national stockpile would simply break the bank. So some public-health experts are calling for an international supply of Tamiflu that could be deployed by the WHO when a pandemic threatens. In unpublished work, Ira Longini, a biostatistician at Emory University in Atlanta, Georgia, has calculated that about 120,000 courses of Tamiflu, if deployed rapidly to treat the sick and protect their families — and if combined with strict quarantine of their houses — could even nip a pandemic in the bud at its point of origin.

Stöhr thinks the idea of ring-fencing outbreaks in this way is "well worth investigating". But Longini's model depends on assumptions about transmissibility and initial death rate that may prove to be wrong. And given the poor infrastructure in many of the Asian countries in which a pandemic virus is most likely to arise, such measures might prove hard to implement in practice. Before embarking on an effort to persuade sceptical governments to invest in such a plan, says Stöhr, there has to be much more confidence in the possibility that it could be made to work.

Uncertainty, unfortunately, is the name of the pandemic flu game. And the problem, for those trying to work out how to organize the first line of defence, is that politicians are averse to spending large sums of money when they don't know the odds — or even whether they'll still be in post when the bet comes in. ■

Alison Abbott is Nature's senior European correspondent.

1. Treanor, J. J. et al. *J. Am. Med. Assoc.* **283**, 1016-1024 (2000).
2. Cooper, N. J. et al. *Br. Med. J.* **326**, 1235-1239 (2003).
3. Zambon, M. & Hayden, F. G. *Antivir. Res.* **49**, 147-156 (2001).
4. Hayden, F. G. *Phil. Trans. R. Soc. Lond. B* **356**, 1877-1884 (2001).
5. Fowler, R. A. et al. *J. Am. Med. Assoc.* **290**, 367-373 (2003).
6. Seto, W. H. et al. *Lancet* **361**, 1519-1520 (2003).



Bitter pill: governments will have to make tough decisions about who should receive Tamiflu.

BUSINESS

Wall Street's gradual green revolution

Companies and investors are starting to wake up to the powerful impact that global warming could have on their bottom line. Emma Marris reports.

What do atmospheric scientists, investment managers, and some of the world's top engineering firms have in common? An imperative to get moving on climate change, a 10 May meeting at the United Nations in New York was told.

Kofi Annan, Al Gore and media mogul-turned-philanthropist Ted Turner told some 400 institutional investors and business leaders that the time had come for investment decisions to hinge on how willing a business is to prepare for climate change.

The day before the UN conference, the iconic US engineering company General Electric (GE) took out six pages of consecutive colour advertising in the newspapers to unveil its new 'green' image. GE pledged to double its investment in environmental research and development to \$1.5 billion a year by 2010 and — rather more cautiously — to cut its own greenhouse-gas emissions by 1% by 2012.

"Real leaders don't wait to be pounded into submission by government process," Paul O'Neill, former treasury secretary under President George Bush and, before that, chief executive of Alcoa, the world's top aluminium producer, told the meeting. Along with other business leaders present, O'Neill argued that international companies should take the leap into cleaner technologies before governments push them there.

Not every investment manager is so sanguine about the commercial promise of these technologies, however. Abby Cohen, a managing director at international investment-banking firm Goldman Sachs, said that investment funds

specializing in 'green' companies had failed to outperform the market.

And people who manage real money on Wall Street are notoriously reluctant to take investment advice based on anything as warm and fuzzy as concerns about the planet's future. So the onus was on the conference organizers to reveal global warming as a cold, hard and expensive fact.

"This is not an environmental conference," said Denise Nappier, state treasurer of Connecticut and co-chair of the meeting. "It's about our economic health. We're learning that they're intertwined — but today, it's all about the money." The meeting was organized by Ceres, a Boston-based coalition of investors and environmentalists.

Many of those present put their names to a 'call to action' organized by the Investor Network on Climate Risk, a group of institutional investors, mostly representing state pension funds and labour unions, who currently manage \$3.2 trillion — about a sixth of all the money managed by institutions on Wall Street.

Its backers hope their collective clout will help to encourage market analysts to factor in the risks of climate change when advising investors. They also want

companies to disclose what their carbon emissions are, what a changing climate



Heated debate: figureheads at a United Nations meeting called for a greener business ethos.

S. B. MARKISZ

might do to their balance sheet, and what they propose to do about it.

The financial risks posed to businesses by climate change include possible costs for upgrading operations in response to limits on carbon emissions in countries that choose to impose them, for adjusting to unpredictable and destructive weather, and for litigation.

Putting the pressure on

Few companies release detailed information on this kind of risk. But shareholders have already pressured some large ones — including Ford, the New York-based merchant bank JP Morgan Chase, and oil giant Chevron — into disclosing such risks in detailed analyses filed with the US Securities and Exchange Commission.

Some observers are uncomfortable with this type of investor pressure on company behaviour. "I'd have more confidence in a chief executive and a board than a group of institutional investors trying to force a company to do things when they don't have all the information," says William O'Keefe, head of the George C. Marshall Institute, a conservative science-policy think-tank based in Washington DC.

But GE's 9 May announcement indicated just how seriously some major corporations are taking the issue. "We're investing in environmentally cleaner technology because we believe it will increase our revenue, our value and our profits," said Jeff Immelt, GE's chief executive, "not because its trendy or moral."

Immelt's sentiment is backed up by an array of major GE customers. John Walsh, for instance, a vice-president of the Canadian



GENERAL ELECTRIC

On the right track? General Electric's locomotives are being redesigned to lessen pollution.



Pacific Railway, says GE's new hybrid-powered Evolution Series locomotives, which his company is buying, will be "the first step in weaning us off fossil fuels".

Welcoming the GE plan at the New York meeting, Al Gore hinted that other major companies may soon follow suit. "There are other shoes about to drop," the former US vice-president said, "some that will surprise people." But despite such optimism, not every major corporation is taking climate change to heart. The one most often seen as hostile to the idea is ExxonMobil. Earlier this month, *Mother Jones* magazine reported that the company was secretly supporting a slew of groups that seek to cast doubt on climate-change science. In New York, Gore branded the Houston-based oil company as "part of the problem".

ExxonMobil was due to hold its own annual shareholders' meeting in Dallas on 25 May. A shareholder resolution there called for, among other things, a report on the company's compliance with the Kyoto Protocol in countries that are signatories to it. The resolution is being backed by Institutional Shareholder Services, an influential group that researches companies and issues recommendations on voting to large investors.

In a company statement, ExxonMobil said it believes that "the scientific evidence on greenhouse-gas emissions remains inconclusive and that studies must continue — while tangible actions are taken to address potential impacts".

GE's stock barely budged after the launch of its green initiative: perhaps investor sentiment was unmoved by Ecomagination, the project's clunky title. But the company remains confident that its new strategy is the right one. "The reception from customers, policy-makers, and even from the public, has been exceptionally strong and positive," says company spokesman Peter O'Toole. ■

IN BRIEF

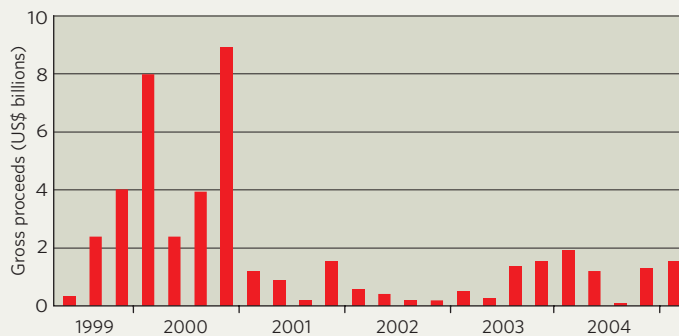
CHEMICAL BOOM US chemical companies are enjoying a powerful and unheralded boom in sales and profits, an industry-wide survey has found. The top 50 producers in the United States saw their total combined sales grow by 23% to \$253.9 billion during 2004, according to an annual survey by *Chemical and Engineering News*. Profits surged dramatically, from \$11 billion to \$18.3 billion. Strong overall demand and high oil prices — reflected in the prices of petrochemical products — are fuelling the boom; petrochemical suppliers such as ExxonMobil Chemicals saw sales growth of almost 40%.

GOING SOFT ON SOFTWARE PATENTS The European Parliament is set to consider a range of amendments that could narrowly confine software patents on the continent. Software-industry representatives fear that the amendments — to a patent directive proposed by the European Commission and agreed by member states — would not adequately protect companies' software innovations. But socialist and green members of parliament who back the proposals — led by Michel Rocard, the former French prime minister — say that less patenting will make it easier for small companies to innovate.

TIME TO MEASURE UP The US National Institute of Standards and Technology (NIST) is asking businesses and scientists to chip in with ideas for the kinds of measurement standards it should establish over the next decade. The agency has appealed for input on the standards that will be needed to measure everything from nanotechnology equipment to broadband communications, in preparation for a strategy it plans to draw up by 2007. "We need to be certain that the US measurement system is robust," says Hratch Semerjian, acting director of NIST.

MARKET WATCH

Proceeds of public stock offerings in biotechnology, by quarter



The first few months of 2005 have deflated investors' hopes for a biotech boom, industry analysts say.

During 2003 and 2004, global investment in biotechnology — including money raised in public share offerings (above), as tracked by publisher BioWorld, based in Atlanta, Georgia — was picking up again after a slump in the previous couple of years.

No one expected a repeat of the money rush that anticipated the human genome sequence. But there had been high hopes for a more modest revival.

"Overall, 2004 was a very good year" for the sector, says Brady Huggett, managing editor of BioWorld. "But things started to taper off as the year came to a close, and that's continued."

So biotechnology firms are finding it tough to raise money on the stock

market right now. Huggett says average initial public offerings in the sector have been bringing in about \$50 million apiece — not enough to sustain a typical research-based company for long.

The generally lacklustre performance of stocks so far this year has made many companies shy of making an initial public offering. And on 28 February, Massachusetts-based Biogen IDEC had to withdraw Tysabri, a treatment for multiple sclerosis, halving the value of the company's own stock and dragging down confidence in the sector as a whole.

But it's the state of the science that dictates market sentiment, Huggett says. And although stem cells and RNA interference show long-term promise, their commercial application remains some way over the horizon. ■

Scientists need back-up by climate organizations

SIR — George Monbiot and the other authors of the Correspondence letter “Time to speak up for climate-change science” (*Nature* **434**, 559; 2005) call on climate scientists to defend the scientific state of knowledge in public debate.

However, more is needed than just to speak out in the media. To spread scientific knowledge effectively, the science community must engage in a dialogue with policy-makers, federal agencies and the private sector.

For single scientists this task may be too time-consuming, on top of their research and teaching activities. Appropriate structures, closely linked to the science community and enjoying high credibility, are needed to support them in this task.

In Switzerland, such structures have been in existence for more than ten years. ProClim (www.proclim.ch), a forum run by the Swiss

Academy of Sciences, and the academy’s Advisory Body on Climate Change (Occc) actively follow the public debate, selecting important issues and working with the science community to prepare assessments (see for example, Extreme Events and Climate Change 2003; www.occch.ch/reports_e.html). ProClim maintains a valuable database of Swiss climate and global-change experts, which allows it to issue position papers and comments for the public debate, reducing the effort for individual scientists.

The work of ProClim and the OccC is widely accepted. Swiss ministers and parliamentary committees regularly ask for direct advice and consult scientists contacted through this office. But this service costs money: it is funded through the Swiss Academy of Sciences and by the Swiss government.

Similar organizations exist in some other countries — for example the German Advisory Council for Global Change (www.wbgu.de) and the Austrian Climate Portal (www.accc.at) — but not everywhere in Europe.

We, and colleagues at other Swiss universities and institutes, consider that such offices, in addition to independent organizations such as www.realclimate.org, provide an efficient way for science to enter into a dialogue with the public and policy-makers.

Christian Körner*, **Heinz Wanner†**,
Christoph Ritz‡

*Institute of Botany, University of Basel, Schönbeinstrasse 6, CH-4056 Basel, Switzerland

†Institute of Geography, University of Bern, Hallerstrasse 12, CH-3012 Bern, Switzerland

‡ProClim, Swiss Academy of Sciences, Schwarztörstrasse 9, CH-3007 Bern, Switzerland

Love of nature led Beuys to new artistic language

SIR — As a geneticist-turned-artist, I enjoyed Martin Kemp’s Science in Culture article on the German artist Joseph Beuys, “The Pied Piper of Düsseldorf” (*Nature* **434**, 141; 2005). However, I believe that Beuys’ relationship with the sciences goes even deeper than portrayed here.

Beuys was always a great lover of animals and plants, and not just as a youth. His expertise in herbs and spices made him a great chef. During the war years, Beuys was close to a superior officer, Heinz Sielmann, also a naturalist. Although Beuys gave up ideas of medical training quite early on, he continued to develop his knowledge of natural history through Sielmann. It is through him that Beuys met Konrad Lorenz, who had established a comparative-ethology department at the Max Planck Institute of Buldern, Westphalia, in 1950.

These contacts gave Beuys a profound respect for nature, both biological and physical. This may be seen in two distinct aspects of his work. One aspect involved a series of specially chosen animals — the hare, the swan, the bee, the stag, the coyote — which existed for Beuys as metaphors for a kind of biological organization that seemed to him to offer a real source of reflection for mankind (as in *Honey Pump in the Workplace*, 1977). Many of the drawings in *The Secret Block for a Secret Person in Ireland* (1936–76) make reference to these animals, which belong, according to Beuys, to both mythology and natural science.

The other aspect of Beuys’ sculptural work invoked the phenomena of heat and

electricity. Beuys did not use these phenomena in a strictly scientific sense, but he went beyond a purely sculptural use of these materials. Thus copper, felt, grease (which could itself melt during an installation) all contribute to the ‘experience’ of the exhibited art-object. Beuys’ idea of potential electric energy, for example, is revealed in the series *Fond VII/2* (1967–84) through piles of cut and layered felt that rise 80 cm to 150 cm above the ground, linked by frail copper wiring that gives the visitor a sense of what stored (potential) energy might ‘look like’.

My own work on a web-based sci-art project called “invisible” (www.invisible-cities.com) leads me to believe that one cannot easily divide Western culture into art on the one hand and science on the other.

Rather than seeing Beuys’ contribution as part of a long history of utopian visions, I personally believe that Beuys developed a language through plastic form which, although not wholly reducible to natural science, was not wholly reducible to art either. As such, Beuys is to be congratulated.

Pete Jeffs

83 rue de Bagnolet, Paris 75020, France

Head of Lorenz Institute is not to blame for delays

SIR — The members of the advisory board of the Konrad Lorenz Institute for Ethology in Vienna were surprised by erroneous statements in your News story “Viennese lab renovations stall as cash goes unspent” (*Nature* **434**, 550; 2005). The reader is left

“My own work on a web-based sci-art project called “invisible” leads me to believe that one cannot easily divide Western culture into art on the one hand and science on the other.” — Pete Jeffs

with the wrong impression that a substantial proportion of the advisory board’s membership is critical both of the work of Dustin Penn as its director and of the allocation of the institute’s finances.

As chair of the board, I feel obliged to clearly state, in the name of the board, that its members fully support Penn and his research plans.

It is correct that renovations and construction of facilities are needed. However, Penn is in no way to be blamed for the delay of these activities, which have turned out to be much more difficult than could have been anticipated.

The board is also aware of unspent money. However, this money has been put aside for investments needed when the new facilities are completed.

Friedrich G. Barth

Department of Neurobiology and Behavioural Sciences, Faculty of Life Sciences, University of Vienna, Institute of Zoology, Althanstrasse 14, A-1090 Vienna, Austria

Nature has learned from several sources that the board of the Konrad Lorenz Institute is not unanimous in its support for the work of its director. Nature therefore stands by the content of this News story — Editor, Nature.

COMMENTARY

Controlling avian flu at the source

Global agricultural authorities should harmonize with the public-health sector to ensure the exchange of flu virus samples, and establish a single international standard for vaccines, say **Robert Webster** and **Diane Hulse**.

We know that a flu pandemic will occur, but not when. Already in Asia the H5N1 avian flu virus has infected at least 88 people, killing 51 of them. For these outbreaks to become a pandemic, the only ingredient lacking is consistent human-to-human transmission. We do not know if the H5N1 vector has this capability — if it does, the effects will be catastrophic.

Certain aspects of pandemic planning have been well considered. But a global strategy for preventing pandemics at their source — in the animals, mostly poultry, that carry the virus — has received relatively little attention. Here we highlight two examples of successful intervention strategies for controlling H5N1 transmission to humans, which we believe merit greater consideration.

In the first documented instance of bird-to-human infection with the H5N1 flu virus in 1997, Hong Kong reacted by destroying its entire poultry population of 1.5 million birds within three days. Many believe that this action averted a pandemic by immediately removing opportunities for further human exposure. In addition to culling, Hong Kong introduced better market conditions, sanitation and vaccination. More recently, Thailand has been successful with the time-honoured agricultural practice of surveillance, restriction of poultry movement, and culling, with farmer compensation.

Both these successes were achieved at high cost, in terms of the number of birds culled and money spent. But moves by international agencies should now help countries achieve the same results with fewer sacrifices.

Access to virus samples

Because of the role of animals in any human pandemic, the United Nations' Food and Agriculture Organization (FAO) and the World Organization for Animal Health (OIE) are involved in pandemic planning, together with the World Health Organization (WHO). The mission of these agencies is in many ways complementary: the WHO protects human health, the OIE protects the health of animals, and the FAO is responsible for animal productivity. But when it comes to pandemic planning for flu, these agencies are not always in harmony.

For example, the lack of animal H5N1 flu viruses collected from the field for analysis is



STR/AFP/GETTY IMAGES

There is now no doubt that the H5N1 avian flu virus is endemic in domestic ducks in Asia.

an area of concern (see *Nature* **435**, 131; 2005). Such analyses are essential for vaccine design and for effective policy decisions. The exchange of animal H5N1 flu viruses between countries and researchers is hampered by free-trade embargoes, national pride, intellectual property, lack of political will, biosafety and bioterrorism considerations, and inadequate infrastructure. There are no simple answers to these and many other difficulties. But such problems are largely man-made, and so should be resolvable.

One of the aftermaths of the severe acute respiratory syndrome (SARS) epidemic is a raised profile of laboratory accidents as sources of disease. Countries with inadequate infrastructure to work with H5N1 influenza are now advised to use molecular probes to detect virulent strains, and not to attempt to isolate H5N1 viruses, which requires biosafety level 3 (BSL3) facilities. Although this seems like sound advice, others may argue for international provision of mobile BSL3 laboratories,

trained staff, and assistance to build up local infrastructure. The FAO has been working diligently in this respect, but the funding is running out. This is one of many reasons why there are so few H5N1 virus samples from humans and other animals

Setting an example

H5N1 bird flu is not a new threat. The 1997 Hong Kong outbreak killed 6 of 18 infected people¹. Indeed, the situation there today owes much to the response since then. Although no H5N1 flu viruses have been isolated from domestic poultry or humans in Hong Kong since 2004, the virus has been circulating in herons and falcons. It is also notable that the H5N1 virus has continued to evolve: in late 2002 it acquired the ability to kill its natural host, wild waterfowl, and spread across ten countries in southeast Asia². The virus has also expanded its host range to include tigers and domestic cats³. The concern is now whether



Even simple measures to curb infection, such as cleaning poultry cages, are not yet widely used.

it will acquire consistent human-to-human transmission.

So what has Hong Kong done to successfully control H5N1 bird flu? The steps taken include: modifying live-poultry marketing practices by banning ducks, geese and quail (the original sources of flu); introducing two 'clean days' per month, when all markets are simultaneously emptied and cleaned; vaccinating all locally raised and imported poultry with an inactivated H5N1 vaccine; and introducing unvaccinated 'sentinel' poultry into each chicken house. These goals have been achieved by model interagency cooperation between the Hong Kong departments of health, agriculture, and food and safety, together with the University of Hong Kong. But why has the Hong Kong model apparently been ignored elsewhere?

The easy answer is that only Hong Kong can afford such rigorous control; the approach is too costly for other, larger countries. But although clean days and improvements to sanitation are disruptive, both are relatively inexpensive. Even simple changes, such as the separation of ducks and chickens in live poultry markets, have not been strongly advocated by international agencies. The question of vaccination is more complex, largely because agricultural vaccines vary in quality.

Although both are efficacious, the flu vaccines for humans and for chickens meet different standards. The few comparative tests that have been done on agricultural vaccines from different suppliers show that some are good and some are bad. Bad vaccines prevent the symptoms of disease but not virus excretion, which can lead to later infection. One of the many arguments against the use of agricultural vaccines is that they promote the selection of mutations in the circulating virus, perpetuating the risk of infection either in the original species or in others. This effect has been documented recently for the H5N2 vaccines used in Central America⁴. However, mutations and selection will occur whether vaccines are used or not. In Hong Kong, sentinel birds provide a way to

monitor virus excretion and the generation of mutations.

We propose that global agricultural authorities harmonize with the human public-health sector to establish a single international standard for a vaccine that is based on antigen content. The current dogma is that agricultural vaccines do not produce sufficient immunity to completely prevent transmission, making culling the preferred alternative. But it may be that high-quality vaccine strains matched to the circulating strains can reduce the level of circulating virus below the transmissible level. The technology for producing inexpensive agricultural vaccines using reverse genetics is available and should be developed.

Harmonizing vaccine use

Arguments about culling and vaccination arise particularly in Thailand and Vietnam. There is now no doubt that highly pathogenic H5N1 avian influenza virus is endemic in domestic ducks in Asia. In Thailand and Vietnam, domestic duck raising increases the likelihood of H5N1 infection by up to eight-fold. But the role of ducks in the continuing spread of H5N1 is complicated because some virus strains kill ducks whereas others are benign in ducks but lethal for chickens and probably humans.

Since December 2004, 41 people have been infected with H5N1 in Vietnam and 16 have died, but none have been infected in Thailand. Why this difference? Thailand has pursued a massive surveillance of poultry in its 'X-ray programme'; trained personnel visited all villages in late 2004 to collect virological and serological samples. The culling of ducks in Thailand (with farmer compensation) has reduced the flocks of ducks that were positive for H5N1 from almost 40% infected in October 2004 to almost undetectable levels in March 2005. Thus, reducing H5N1 infection in poultry clearly reduces the threat to humans.

Could vaccinating ducks have similar benefits to culling? Duck vaccines for flu are in an even worse state than chicken vaccines; virtu-

ally nothing is known. Unless high-quality vaccines are developed and used, H5N1 viruses that are benign in ducks but lethal for chickens and humans will be selected (although this is likely to occur naturally anyway). Our aim must be to reduce the virus load in ducks below the transmissible level, as was achieved by vaccinating chickens in Hong Kong. This is probably the approach that has the best chance of reducing the inevitability of H5N1 acquiring consistent human-to-human transmission.

Thai authorities know that even if H5N1 is eradicated from domestic poultry this year, the virus will probably return from neighbouring countries. Also, it is not known whether highly pathogenic H5N1 variants are endemic in the wild waterfowl that breed in Siberia and migrate through the region. Flu, like SARS, knows no borders, and it is likely that H5N1 will continue to cross them. Consequently, relying on culling alone is a risky strategy.

China, which does not export much poultry, adopted the generalized use of poultry vaccines to control the H5N1 epidemic in 2004, and has reported no clinical outbreaks in domestic poultry since. Indonesia has also used poultry vaccines, although there have been reported cases of H5N1 in poultry and pigs this year. Vietnam has recently decided to start testing poultry vaccines this summer.

Thailand is a major poultry-exporting country. And its trading partners would prefer it not to use poultry vaccines, which can mask residual H5N1 viruses. However, Thailand is now investigating the use of flu vaccines for backyard poultry, free-range ducks and fighting cocks (not commercial poultry). This decision represents a major shift in policy. The use of H5N1 vaccines for such 'open range' poultry is a prudent step, and should be encouraged in other countries in the region.

Commercial poultry and backyard poultry are currently considered to be linked. Therefore during an outbreak all birds are culled, which is expensive. The OIE is now considering rule changes that would recognize different categories of poultry (closed chicken farms and backyard poultry). Countries could then move to having a system of zones based on poultry populations with different disease status. In this way, they could develop disease-free zones, perhaps by the selective use of vaccines, and recover some export capacity. We hope that the OIE's initiative indicates growing harmony among international agencies. ■

Robert Webster and Diane Hulse are in the Department of Infectious Diseases, St Jude Children's Research Hospital, Division of Virology, Memphis, Tennessee 38105, USA.

1. de Jong, J. C., Claas, E. C., Osterhaus, A. D., Webster, R. G., & Lim, W. L. *Nature* **389**, 554 (1997).
2. Li, K. S. et al. *Nature* **430**, 209–213 (2004).
3. Kuiken, T. et al. *Science* **306**, 241 (2004).
4. Lee, C. W. et al. *J. Virol.* **78**, 8372–8381 (2004).

N. BEHRING-CHISHOLM/GETTY IMAGES

Time	Flight	Destination	Gate	Status
16:15	CX 872	San Francisco	1	Boarding
	AA 6118			Cancelled
16:15	JL 706	Nagoya		Cancelled
16:20	CX 828	Vancouver		Cancelled
		Toronto		Cancelled
16:25	CX 731	Dubai		Cancelled
16:25	CX 838	Vancouver		Cancelled
16:30	CX 903	Manila	28	
16:30	OX 2211	Phuket	27	
		Bangkok		Cancelled
16:35	CX 737	Singapore		Cancelled
16:40	BR 856	Taipei	30	
16:45	KA 808	Shanghai/Pudong	18	
16:50	GA 859	Singapore	19	
		Jakarta		
16:55	CA 110	Beijing	17	
17:00	Z 3038	Changsha	34	
17:05	614	Taipei	25	
17:10	703	Bangkok	32	
		Karachi		
17:15	482	Taipei	26	
18:30	SQ 809	Vancouver		
18:40	AC 008	Toronto		
18:50	TG 633	Bangkok		Cancelled
19:05	KA 430	Kaohsiung		Cancelled
19:10	CX 111	Sydney		Cancelled
19:10	CX 135	Melbourne		Cancelled
19:15	MU 536	Shanghai/Pudong	16	
19:20	SQ 869	Singapore		Cancelled
19:35	QF 088	Melbourne		Cancelled
19:40	5J 119	Manila	31	
19:40	CX 468	Taipei	61	
19:40	CX 913	Manila		Cancelled
19:45	CI 642	Taipei	1	
19:55	UA 805	Singapore		Cancelled
20:00	CX 715	Singapore		Cancelled
20:10	QF 086	Brisbane		Cancelled
		Sydney		
20:15	TG 630	Taipei		Cancelled
20:25	CX 107	Auckland		Cancelled
20:40	SQ 010	Las Vegas		Cancelled
20:45	TG 607	Bangkok		

A weapon the world needs

Both bottom-up and top-down planning is needed to prevent a global economic disaster.

Michael T. Osterholm calls for action at all levels.

Influenza experts are more worried than ever about the next pandemic. It could be caused by H5N1, the avian flu strain of such concern in Asia; it could even rival the devastation of the 1918 Spanish flu pandemic. Whatever form it takes, it is sobering to realize that when the last pandemic emerged in 1968, in China, the nation's human population was 790 million and the poultry population 12.3 million; today those numbers are 1.3 billion and 13 billion, respectively. Similar changes have occurred in other Asian countries, creating an incredible mixing vessel for viruses. A pandemic could be unleashed tomorrow or in ten years from now, but the scene for a potential catastrophe is already set.

Every year, seasonal influenza A kills up to 1.5 million people around the world as the disease migrates between the Northern and Southern Hemispheres¹. Current efforts to reduce this global death toll largely involve the delivery of roughly 250 million to 300 million doses of influenza vaccine to the most vulnerable residents in a dozen or so industrial nations. Those fortunate enough to receive vaccines represent less than 5% of the world's current population of 6.5 billion people.

Vaccination is the only meaningful weapon to combat the next pandemic. But the egg-

produced influenza vaccine is based primarily on 1950s technology, which means production could not be immediately ramped up if a global flu pandemic became reality. The 20-year struggle by public-health officials in industrial nations to increase private-sector vaccine production for seasonal flu epidemics is a tale of 'two steps forward and one step back'. Overall, our poor vaccine manufacturing infrastructure, together with national and international failures to push for universal vaccinations, has left the masses vulnerable to annual flu and, worse, the next flu pandemic.

Plan for action

What should we do? The World Health Organization (WHO) recently issued a revised global influenza preparedness plan that provides recommendations for national measures before and during a pandemic². It is meant to help countries develop or update their national flu preparedness plans. The WHO's plan primarily targets public-health officials, although an executive summary has been produced for senior policy-makers who may have a public-health background. Although the WHO's document is a helpful tool for those responsible for preparedness at

the national level, it falls far short of what is needed at either the ground level (local day-to-day planning) or at the international level (long-term planning). But it is critical to consider all three perspectives when planning preparedness and response strategies. Unfortunately, all other national and state plans suffer from similar failings.

The WHO's national-level recommendations are non-specific in nature. For example, the 'Phase 2 health system response'² advises that countries "verify availability and distribution procedures for personal protective equipment and antivirals, and for vaccine, for the protection of persons at occupational risk"; and that they "consider measures to implement". But such recommendations assume that protective equipment, vaccines and antivirals will already be available. They also do not address the difficult ethical questions about allocation: who should get priority for receiving these potentially life-saving products if they exist in limited supply?

Such issues are central to the international and ground-level perspectives, both of which have been neglected in planning for pandemic flu in most countries. If we are really serious about preparing for a flu pandemic, regardless of whether it begins tonight, next

year or ten years from now, we must flesh out and act upon the critical needs at these other levels.

At the international level we must put the availability of a pandemic vaccine at the top of the list. Other priorities include making effective antivirals and protective masks available. Unfortunately, most industrial countries are looking at the vaccine issue through myopic lenses. The primary question seems to be: how do we get enough vaccine in the first months of the pandemic to protect our citizens?

For a classic public-health approach this perspective makes sense. Countries such as the United States, which have aggressive influenza vaccine research programmes, are to be commended. However, a purely national approach fails to consider the nature of the modern world — a world of globally distributed just-in-time inventories for almost all consumer products, including medical supplies. The world today is much more vulnerable to the collapse of trade than it was in 1918.

Wake-up call

The arrival of pandemic flu will trigger a reaction that will change the world overnight. There will be an immediate response from leaders to stop the virus entering their countries by greatly reducing and even ending foreign travel and trade — as was seen in parts of Asia in response to the severe acute respiratory syndrome (SARS) epidemic. These efforts are doomed to fail given the infectiousness of the virus and the volume of illegal crossings that occur at most borders. But government officials will feel compelled to do something to demonstrate leadership. Individual communities will also want to bar 'outsiders'. Global, national and regional economies will come to an abrupt halt.

A vaccine against the pandemic strain produced using current technology would not be available for at least six months after the pandemic starts. And even then, the supply would only be large enough to vaccinate 14% of the global population¹. Our limited stockpiles of antiviral drugs and medical ventilators will seem as inadequate as they did in 1918.

But the problem is not just one of death and disease; trade and economic dependency are also at risk. The global economy has never been measurably threatened by human immunodeficiency virus (HIV), malaria or tuberculosis despite the dramatic impact of these diseases on developing-world populations, particularly sub-Saharan Africa. The global panic created by flu will be different. Today, we have virtually no surge capacity for any consumer product or medical service that might be needed during the 12 to 36 months of a pandemic.

We must demand nothing less than an international effort to develop a new type of influenza vaccine that can be manufactured on a much shorter timescale. This global



The current practice of using eggs to make flu vaccine is too slow to cope with a pandemic.

vaccine will require a new method of production, surge capacity for crises and a detailed plan for distribution. One possibility is to move away from strain-specific vaccines towards generic vaccines that can respond to all virus strains.

A vaccine cannot be delivered fast enough to prevent a virus spreading in those countries where the pandemic first emerges. But by vaccinating people in many more countries we could minimize its impact: imagine if only 2% of the global population became infected instead of 50% (as is likely now). The world would recover much more quickly from the first pandemic wave, and be in a much better position to deal with any subsequent waves of infection.

So, although we cannot prevent a pandemic from happening, we might be able to change its course if we start acting now, and if the pandemic is still a few years away. But if industrial countries continue to develop vaccines for just themselves, they, and everyone else, will remain vulnerable to a global economic disaster. Even if nations vaccinate their entire populations, they cannot remain isolated from a pandemic shock.

Who should be taking the lead on the vaccine issue? We need bold leadership from the group of seven industrialized nations plus Russia (G8) and other developed-world governments. When the G8 leaders next meet in Scotland in July, avian flu will be on the agenda, but major commitments are unlikely. This is not good enough. These nations urgently need to recognize the economic, and security and health threat that the next flu pandemic poses, and invest accordingly.

As well as waking up to reality at the international level, we must also struggle with difficult issues at ground level. For example, there

are no detailed plans in place to staff or equip temporary hospitals, which might be installed in high-school gymnasiums or community centres for as long as one to two years. Nor are there detailed plans on how to handle the dead bodies whose numbers will soon outstrip our ability to process them.

Ethical questions also need to be tackled now, in a public forum. Who will get the extremely limited antiviral drugs that will be available? Any priority setting during the crisis will provoke further dissent and disruption. Both government-sponsored and private health-care delivery systems have conducted little planning around this issue.

We also cannot predict how many health-care workers will continue to place themselves at high risk of infection by taking care of influenza patients, if vaccine and protective equipment are not available. Health-care workers will become ill and die as quickly as the rest of the public, or even faster, particularly if they have limited protective equipment.

It is essential to think now about the possible use of lay volunteers in hospitals — especially of those who survive the first wave of infection. Such survivors may have gained immunity before a vaccine has become available, and may want to assist clinicians. The strong medical arguments against using lay volunteers — grounded in both liability concerns and professional hubris — must be addressed.

Time's up

Time is running out to prepare for the next pandemic. There is a critical need for comprehensive medical and non-medical pandemic planning at the ground level (involving many in the private sector), that goes beyond what has been considered so far. National, regional or local plans based on general statements of intent or action will be meaningless in the face of a pandemic. Specific operating blueprints to get through 12 to 36 months of a pandemic are essential. For example, determining how food might be supplied to local populations when transportation and food-processing plants shut down will require a level of planning not yet included in any national or regional plans.

At the international level, world leaders need urgently to consider what they can do now. When the pandemic hits close to home, leaders will do their best to react and cope. But real leadership, particularly by the G8, means making tough decisions now. We must act with decisiveness and purpose if we are to create a pandemic vaccine that has a chance of making a real difference.

Michael T. Osterholm is at the Center for Infectious Disease Research and Policy, University of Minnesota, Academic Health Center, 420 Delaware St SE, MMC 263, Minneapolis, Minnesota 55455, USA.

- Osterholm, M. T. *N. Engl. J. Med.* **352**, 1839–1842 (2005).
- http://www.who.int/csr/resources/publications/influenza/WHO_CDS_CSR_GIP_2005_5/en

Global task force for influenza

Early detection and rapid response to bird flu, on a global scale, will drastically cut the costs of dealing with a full-blown human flu pandemic, argue **Ron Fouchier, Thijs Kuiken, Guus Rimmelzwaan and Albert Osterhaus.**

A human flu pandemic could cause 20% of the world's population to become ill. Within a few months, close to 30 million people would need to be hospitalized, a quarter of whom would die¹. Although these estimates are speculative, they are among the more optimistic predictions of how the next flu pandemic might unfold.

Like most emerging virus infections that threaten human health, flu outbreaks originate from animal reservoirs. Because of rapidly changing human behaviour and animal ecology², infections are spreading faster and farther. Patchy research on these outbreaks, and poor coordination between different disciplines in response to them, are limiting our ability to deal adequately with the threat they pose to human health. We propose establishing a permanent global task force to control a flu pandemic, in which relevant agencies would work together with leading research groups from different disciplines.

Close watch

An outbreak of avian flu among chickens in the Netherlands in 2003 led to the culling of 31 million birds. Initially, the ministry of health issued warnings, but the ministry of agriculture insisted that there was no risk to human health. "I've worked with sick chickens all my life and never become ill before," was the sentiment commonly voiced by veterinarians involved in the culling activities. But by the end of the outbreak, 89 people were confirmed to be infected with the flu virus strain H7N7; one died and the rest experienced eye disease or common flu symptoms³.

In southeast Asia, several countries are currently affected by the H5N1 strain of flu virus. Animal health authorities in some countries have delayed reporting the disease in poultry, in some cases by several weeks, and this has contributed to the failure to contain these epidemics. Such delays have important implications for human health; most human cases of avian flu to date have been linked to contact with infected poultry.

Traditionally, the poultry industry has approached avian flu as an economic problem. Poultry are tested only when animals become sick and die, and even then testing is generally restricted to a search for highly pathogenic strains of the H5 and H7 flu virus subtypes, excluding the other 14 subtypes and less pathogenic strains. Yet the past three human flu pandemics did not originate from H5 or H7 viruses⁴. Moreover, the reporting of viral infections is often delayed because of authorities' conflicting mandates to report disease



Pulling together: agency cooperation is essential to prepare for the threat posed by avian flu.

occurrences and to protect their country's export status for animal products.

At the international level, trade regulations for poultry and poultry products are defined by the World Organization for Animal Health (OIE) based in Paris. However, there is geographical variation in production and distribution systems for poultry (for example, farms may mix species, or have open access to wild birds; birds may be taken live to market or killed beforehand). The extent to which affected countries follow effective outbreak-containment measures is also variable; culling and vaccine-deployment strategies are not standardized. And measures to protect poultry workers during flu outbreaks vary greatly from country to country.

This lack of international harmony in detecting and dealing with avian flu extends to the assessment of human infections. On the basis of 51 deaths out of 88 laboratory-confirmed cases⁵, the fatality rate from H5N1 flu virus infection in humans is estimated to be around 60%. But the true incidence of infection, and the associated spectrum of disease symptoms, may be very different⁶: data on post-mortem investigations are largely lacking, for example, limiting our knowledge of which tissues the virus attacks and how it causes disease. It is therefore impossible to determine which of the available and very variable (in terms of their pathological outcome) animal models — mouse, ferret, cat, pig, macaque — most closely resembles the disease in humans.

Specific monitoring of virological, serological and clinical parameters is urgently needed for people at risk. Also needed are detailed autopsies to characterize the disease, and the subsequent establishment of appropriate animal models to evaluate available intervention strategies. For poultry, bird populations should be actively surveyed for all subtypes of flu virus, using high-throughput technology; production and distribution systems should be modified; and stricter adherence to containment measures achieved when there is an outbreak.

Integration

To obtain a better global picture of the threat posed by avian flu, it is imperative to investigate the virus in wild bird populations. Wild birds, particularly migratory ducks, geese and shorebirds, are the natural reservoir of influenza A viruses, which can infect other avian and mammalian species⁷. But information about flu in wild birds is still limited. A widespread and integrated approach is needed to understand the dynamics, epidemiology and pathogenesis of these virus infections in wild birds, and the potential routes of virus transmission. For example, transmission from wild birds to poultry and mammalian species, including humans, may result from direct contact with production systems, the wild bird trade, and smuggling.

To limit the effects of flu on public health and livestock production, integrated and effective action from all the disciplines



Group effort: to tackle flu effectively, greater coordination is needed between the health workers providing vaccinations (above) and the veterinarians monitoring poultry (below).

involved is urgently needed, rather than ad-hoc responses at the national level. To this end, we advocate a global flu task force as an essential element of the World Health Organization's (WHO's) flu pandemic preparedness plan.

Task-force agenda

This task force should consist of leading specialists in the fields of human and animal medicine, virology, epidemiology, pathology, ecology and agriculture, as well as experts in translating science into policy. The task force must be able to respond rapidly and effectively, so data must be exchanged and integrated quickly as they emerge. Thus, when the need occurs, outbreak-management teams can be formed and targeted at a specific outbreak in a defined area of the world. These teams should consist of task-force representatives as well as local experts and policy-makers from the countries affected.

The viability of such an approach is clearly illustrated by the WHO teams formed to deal with the severe acute respiratory syndrome (SARS) outbreak⁸. Although the achievements of these teams were impressive, the integration between their activities could have been better, and better use could have been made of pre-existing human and animal health networks.

For the current H5N1 virus outbreaks, similar WHO teams have been established, but they probably suffer from the same limitations. Given the large geographical area in which the H5N1 virus has become endemic, and the greater potential for rapid virus spread, an efficient, effective, outbreak-management team strategy, with centralized guidance, is urgently needed. As key international players related to human and animal health, the WHO, the OIE and the United Nations Food and Agriculture Organization (FAO) are

in the unique position to provide the political goodwill and to endorse the proposed integrated approach to the problem of avian flu, which does not respect national borders.

The immediate duties of our proposed task force are fourfold. First, to gain insight into the global picture of flu, taking into account temporal and geographical variation of the virus, in the different species involved (wild birds, poultry, humans, other domestic animals such as pigs, horses and cats, and other wild animals such as seals, cetaceans and tigers). Second, to prioritize research and integrate knowledge of different disciplines on influenza virus infections. Third, to advance intervention strategies for animal outbreaks and human cases. And fourth, to translate knowledge into policy advice, emphasizing the integration of human and animal health strategies.

In dealing with a specific virus outbreak in a defined geographical area, the task force's

outbreak teams would focus on: geographical distribution and the species involved in the outbreak; gaining detailed knowledge of the different aspects of the viruses and their interactions with hosts; assessing the risk of spread; and advising on the best options for intervention.

Cost-effectiveness

Early detection and rapid response to avian flu at the global level will greatly reduce the direct and indirect costs of dealing with a full-blown flu outbreak. For example, in the H7N7 outbreak in the Netherlands and the H5N1 outbreaks in Thailand and Vietnam in 2003, the agricultural costs alone were estimated to be US\$348 million, US\$880 million and US\$120 million, respectively. These costs do not take into account the costs of human sickness and death, or the damage done to other areas of the economy, such as tourism.

By contrast, we estimate that the cost of setting up and operating a global task force would be less than US\$1.5 million a year. The costs of a global human flu pandemic would far exceed the costs estimated to have resulted from the 2003 SARS pandemic, and dwarf the amount of money needed for effective containment and prevention.

Ron Fouchier, Thijs Kuiken, Guus Rimmelzwaan and Albert Osterhaus are at the National Influenza Centre and Department of Virology, Erasmus Medical Centre Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands.

1. Stohr, K. & Esvel, M. *Science* **306**, 2195–2196 (2004).
2. Kuiken, T., Fouchier, R. A., Rimmelzwaan, G. F. & Osterhaus, A. D. *Curr. Opin. Biotechnol.* **14**, 641–646 (2003).
3. Fouchier, R. A. et al. *Proc. Natl Acad. Sci. USA* **101**, 1356–1361 (2004).
4. Fouchier, R. A., Rimmelzwaan, G. F., Kuiken, T. & Osterhaus, A. D. *Curr. Opin. Infect. Dis.* **18**, 141–146 (2005).
5. www.who.int/csr/disease/avian_influenza/country/en
6. de Jong, M. D. et al. *N. Engl. J. Med.* **352**, 686–691 (2005).
7. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. *Microbiol. Rev.* **56**, 152–179 (1992).
8. WHO *Lancet* **361**, 1730–1733 (2003).





C.S. XINHUA/AP

Is China prepared for microbial threats?

There is no bigger acute microbial threat to China, and to the rest of the world, than an influenza pandemic, and no better time to prepare for this eventuality than now. **David Ho** asks what more China could be doing.

Severe acute respiratory syndrome (SARS) was unquestionably a loud wake-up call for China. The outbreak of this transmissible disease caught the health authorities unprepared. The country's disease-surveillance system was slow to recognize and report the new syndrome, and the magnitude of the threat was not, at first, fully appreciated. The problem was exacerbated because health officials did not, initially, handle the burgeoning epidemic with the transparency that the crisis required. A chance to nip the outbreak in the bud was lost, and the consequences were devastating for both China and the rest of the world.

SARS spread rapidly in early 2003 from the southern province of Guangdong to other parts of the country, most notably Beijing and Hong Kong. It also spread by air travel, quickly surfacing in places such as Vietnam, Singapore, Taiwan and Canada. SARS was recognized as a new disease by medical professionals in Hong Kong and international health officials working in Vietnam. With unprecedented speed, the causative agent was isolated and identified as a previously unknown form of coronavirus.

Fortunately, China took stock of the situation and by April 2003 had made a course correction. In the end, the political will to end the epidemic was impressive, and the new openness in addressing the disease was refreshing. By launching a concerted public-health effort that included draconian measures such as quarantine, China brought SARS

under control faster than anyone could have predicted — the outbreak was over by that summer. The success was not only commendable, but a testament to Chinese resourcefulness once a clear path is apparent. But China — and the world — paid a high price: cities were paralysed for months, billions of dollars were lost, more than 8,000 people were infected, and about 800 lives taken.

High alert

China is now confident that it could handle the re-emergence of SARS. But is it sufficiently prepared for other major microbial threats that will emerge or expand in the near future? The Chinese epidemic surveillance system seems appropriate on paper, with its extensive network of disease centres at the national, provincial and local levels. But many of the deficiencies highlighted during the SARS outbreak remain.

Despite a significant infusion of funds in 2003, the disease-surveillance system is still grossly underfunded, and consequently lacks sufficient human resources and technical capacity. Sufficient resources must be allocated to train enough professionals to safeguard against epidemic diseases. Currently, too few of China's health workers have been properly trained to carry out the task. It is imperative that a cadre of medical and public-health officers are formally trained in special programmes, such as those at the Epidemiologic Intelligence Service of the US Centers for Disease Control or at the World Health

Organization's office in Lyon. Trained individuals could, in turn, mentor others and so build the critical mass of professionals necessary to rejuvenate the Chinese disease-surveillance network.

China's disease-alert system must also be upgraded. The country's prowess in information technology should be applied to create a rapid automated mechanism to report clusters of unusual syndromes or specific infections from its vast network of clinics and hospitals. Such a system should also be alerted when certain pathogenic microbes are identified in the clinical laboratories. The ProMED-mail list for tracking infectious diseases worldwide could serve as a model.

Once the information is received centrally at the Chinese Centre for Disease Control, it must be processed by disease experts who can respond in a thorough and expeditious manner. Anything of concern should be rapidly disseminated to physicians and public-health specialists so that they are on the alert for similar illnesses or infections. When appropriate, such warnings should also go to government leaders, the media and international health agencies. To avert or blunt another catastrophe like SARS, China urgently needs a more efficient and transparent instrument to spot early disease outbreaks and to propagate the warning.

Two fundamental problems must also be addressed to improve China's preparedness for significant microbial threats. First, China must make microbial threats to health a top

priority in its national research agenda. Greater financial resources should be given to microbiologists in academic institutions and government research units. This sort of commitment will undoubtedly result in a higher level of technical competence, which will be valuable during any serious microbial outbreak. A larger critical mass of experts could provide strategic advice and diagnostic tools to fight an epidemic. It could also help develop the therapies and vaccines needed to end the threat. Second, healthcare infrastructure has suffered years of neglect. Among the adverse consequences is a relatively low level of proficiency among physicians in China. This deficit will in turn allow certain epidemics to escape early detection, because these doctors are the first line of defence.

Latest threats

There is no bigger acute microbial threat to China than the next influenza pandemic. The world suffered greatly from three such pandemics in the twentieth century, including the Spanish flu of 1918–19 that killed more than 20 million people. Another flu pandemic is not only inevitable, but overdue. There is no better time to prepare for this eventuality than now, as an aggressive influenza virus, H5N1, has been continuously plaguing avian species over the past two years in about a dozen Asian countries, including China.

This viral variant is deadly in chickens and ducks; it also killed about 40 humans who were infected through direct contact with avian species. Fortunately, this virus has yet to acquire the ability to transmit from person to person (with a few exceptions); otherwise, we would be facing a pandemic already. Nevertheless, the current situation is of grave concern. Chinese scientists have already isolated H5N1 from nasal secretions of pigs, a

species that can replicate both avian and human influenza viruses. It is only a matter of time before H5N1 and a human influenza virus mix and 'reassort' in porcine hosts to yield a new virulent strain that is capable of human-to-human transmission. It seems as though all the essential elements are now in place to produce a new influenza pandemic, most likely emerging from southeast Asia or China.

The world, China included, must respond as if the next pandemic is imminent. In August 2004, the US government released its national pandemic preparedness plan, broadly outlining the actions that will be taken. Various deficiencies were emphasized, including difficulties in implementing certain public-health measures such as quarantine. Despite its technical superiority, the United States faces a major shortfall in manufacturing the right vaccine and in stockpiling anti-influenza drugs. An estimated 89,000 to 207,000 Americans will die in the next pandemic. What would the death toll be in China?

China's recent SARS-fighting experience will give its pandemic response an edge. But this advantage will be offset by a number of factors: China is likely to be hit first or early by the pandemic; its disease-surveillance system and overall healthcare infrastructure are inadequate; the health authorities have yet to come up with a detailed strategic preparedness plan; and it has limited technical resources to produce enough vaccines and drugs to combat the pandemic. There is little doubt that China will be in deep trouble if the flu pandemic were to strike in the next few years. It has a moral obligation to its own people, and to the world, to rectify the situation as soon as possible.

In the aftermath of SARS, China saw many of its health challenges in a fresh light. Prominent among them was another

epidemic transmissible disease. HIV/AIDS reached China in the mid-1980s, but the significant spread began in 1989 among injecting drug users in the southwest province of Yunnan. Following the tracks of heroin, this deadly virus spread to the adjacent provinces of Guangxi and Sichuan, and to the north-west autonomous region of Xinjiang. In addition, a separate HIV/AIDS epidemic exploded in the 1990s in Henan and neighbouring provinces, where unsanitary blood collection led to the infection of countless poor farmers.

Today, China has nearly 1 million individuals who are HIV-positive. But nine out of ten of them do not know that they are infected. About half the Chinese population knows little or nothing about HIV. The Joint United Nations Programme on HIV/AIDS, UNAIDS, has highlighted this epidemic as 'China's titanic peril', suggesting that 10 million people could become infected by the end of this decade. And yet until relatively recently, the Chinese response to the issue was somewhat muted.

The experience of the SARS outbreak has helped to change that. Slowly but surely, China is adopting its SARS-fighting attitude to combat HIV. There is now a centrally coordinated AIDS office with sufficient political will and clout to address the epidemic.

Although much could still be improved, China's recent AIDS policies are moving in the right direction. For example, the government has begun to offer free HIV testing and antiretroviral therapy, and has launched large-scale surveillance in several provinces. Prevention programmes such as condom promotion and distribution, methadone maintenance, and even needle and syringe exchanges are emerging. Similar forward-looking approaches should be applied to a flu pandemic preparedness plan.

Prospects

In the past two decades, the world has been awed by China's single-minded determination to become a nation that is technologically advanced and economically strong. Its economic growth during this period has been unprecedented. Its booming manufacturing capability is beginning to lift the country out of the developing world. It is high time for China to place greater importance on one of the hidden costs of its economic rise: the health of its population. Ultimately, this vital priority will only be attained if the national epidemic surveillance system is significantly upgraded. More emerging infections are inevitable, as the past century attests. China must be better equipped to deal with potential calamities posed by serious microbial threats. After all, there will be no prosperity if there is no health.

David Ho is at the Aaron Diamond AIDS Research Center, Rockefeller University, 455 First Avenue, New York, New York 10016, USA.



The next threat: a chicken is tested for avian flu at a poultry farm in Beijing.



Race against time

A committed, transparent research effort into the detection, prevention and treatment of bird flu is now critical. **Anthony S. Fauci** presents the questions that need answers.

The emergence of the highly pathogenic H5N1 influenza A virus in southeast Asia is a grim reminder of the deadly toll of flu pandemics throughout the ages. This virus has the potential to trigger the next pandemic, which, judging from history, is well overdue. We cannot predict exactly when a pandemic will occur, nor can we know for certain whether the culprit will be H5N1 or a related virus. The only certainty is that it will present extraordinary challenges. Many individuals, research and health organizations, and countries have delineated surveillance goals, are sharing information and reagents, and are asking the multiple research and logistical questions needed to develop an appropriate response. Clearly, there is much to be accomplished, and time is of the essence.

To prepare for and minimize the impact of avian viruses with pandemic potential, a multifaceted approach is essential. This includes a robust, sustained commitment to the research needed to develop new tools and strategies for influenza detection, surveillance, prevention and therapy.

Preparedness for pandemic flu is a global endeavour, with the World Health Organization (WHO) playing a pivotal role. The US contribution to the global effort involves several government agencies that operate in partnership with private industry and academic institutions. One of the main agencies, the Department of Health and Human Services (HHS), will spend about \$419 million in its

Influenza Preparedness and Response Plan for the fiscal year 2005. Within the HHS, the Centers for Disease Control and Prevention (CDC) has the key role of mediating disease surveillance and the public-health response. In addition, the National Institutes of Health (NIH) conducts basic research on viral pathogenicity and molecular evolution. The NIH also carries out research to identify drug targets and new technologies for vaccine production, as well as clinical research to test the safety and efficacy of new diagnostics, vaccines and antivirals (see www2.niaid.nih.gov/Newsroom/FocusOn/Flu04). The NIH's flu research budget for 2005 is approximately \$119 million; it supports work that complements the efforts of the CDC and the Food and Drug Administration (FDA).

Watch and learn

A timely response to a potential flu pandemic requires diligent surveillance as well as an understanding of how flu viruses evolve, spread and cause disease. The NIH Influenza Genome Sequencing Project aims to provide complete sequence data for selected human and avian influenza isolates. When the project began in November 2004, only seven human influenza H3N2 isolates had been completely sequenced and deposited in the NIH's GenBank database. Six months later, the sequences of 150 human H3N2 isolates had been determined and made available. Information now deposited in GenBank from various sources includes sequences of 110 complete H5N1

genomes, 4 complete H2N2 genomes, 4 complete H7N3 genomes and 30 complete H9N2 genomes (see www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). These data are invaluable for tracking viral evolution and transmission, and for developing new drugs, vaccines and diagnostics.

Adding to our understanding of the molecular mechanisms driving flu virus emergence in animals and their spread to humans is a project carried out by investigators from St. Jude Children's Research Hospital, Tennessee. These researchers traced the current H5N1 outbreak in Asia back through a series of genetic reassortment events to a human outbreak that occurred in Hong Kong in 1997 (ref. 1). The genetic reassortments resulted in a dominant H5N1 genotype infecting chickens and ducks in 2003 and 2004. Domestic ducks in southern China were central players in the generation and maintenance of the virus, with migratory birds possibly contributing to the spread of the virus throughout southeast Asia.

It is crucial to continue surveillance studies of this type in order to monitor the spread of flu viruses from animal to animal, animal to human and from human to human. If a virus acquires the ability to transmit efficiently among humans, a pandemic is likely, whereupon aggressive intervention measures will be needed.

Vaccination and the use of antivirals are two of the most important response measures to a flu pandemic. H5N1 viruses are sensitive to the neuraminidase inhibitors oseltamivir (an oral

drug) and zanamavir (an inhalable powder). The HHS and the CDC have created an initial stockpile of 2.3 million treatment courses of oseltamivir, to which more doses will be added. More research and discussion are needed to prioritize the distribution of stockpiled anti-influenza drugs in case of a pandemic.

The NIH currently funds projects to develop and test other flu inhibitors, including CS-8958, a long-acting next-generation neuraminidase inhibitor. Other studies are directed towards identifying and evaluating new viral drug targets.

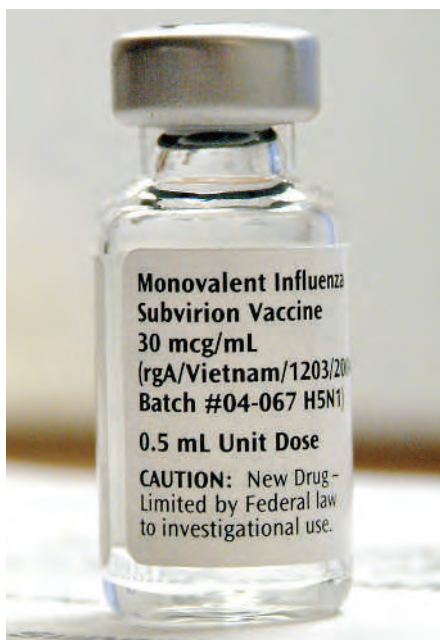
Ramping up vaccines

Controlling an emerging flu pandemic will require a rapid and aggressive response: considerable efforts are being channelled into finding ways to streamline vaccine manufacture to allow for greater flexibility and surge capacity. The first step in vaccine preparation requires the generation of a 'reference virus' — generally a time-consuming process. A new technique called reverse genetics can be used to generate, relatively quickly, reference viruses that precisely match a target flu strain². Using this technique, an H5N1 reference virus was produced within weeks and used to develop an inactivated H5N1 vaccine. A phase I clinical trial is currently under way to test the safety, immunogenicity (ability to stimulate an immune response), and appropriate dosage of an inactivated H5N1 vaccine manufactured by Sanofi Pasteur. The first phase of the trial will enroll 450 healthy adults, aged 18 to 64 years, at three sites in the United States. If the vaccine is shown to be safe, it will be tested in the elderly and in children (see www2.niaid.nih.gov/Newsroom/FocusOn/Flu04/).

The HHS has awarded a contract to Sanofi Pasteur to produce 2 million doses of inactivated H5N1 vaccine to create a stockpile as part of the HHS influenza preparedness plan. This effort will ensure that, should the need arise, the manufacturing techniques, procedures, and conditions for large-scale production are already in place. The movement to large-scale vaccine production in parallel with clinical trials indicates the urgency with which H5N1 vaccine development must be addressed. Waiting for the results of initial trials is normal procedure, but this would significantly delay production of large quantities of vaccine that might be needed to vaccinate health workers, researchers, and possibly the public, in affected areas.

To further broaden the pandemic vaccine portfolio, NIH has initiated the production and clinical testing of H9N2 candidates. The H9N2 influenza virus, although not highly pathogenic to humans, has been the most prevalent subtype of avian flu circulating among birds in Hong Kong and China³, infecting at least eight individuals in the region. The results of phase I safety trials to evaluate the H9N2 vaccine are expected this year.

Immune responses generated by live attenuated vaccines develop more quickly and are



Protective potion: research into vaccines is urgently needed to deal with a flu pandemic.

more robust than those triggered by inactivated vaccines, such as those described above; they also tend to be more cross-protective against related variants of the same virus⁴. Live-attenuated H9N2 and H5N1 vaccine candidates have been generated and their safety, infectivity and immunogenicity will be tested in phase I clinical trials, probably later this year.

All licensed flu vaccines rely on the traditional method of vaccine production: the virus is grown inside chicken eggs. But this method does not readily allow for adjustments to increase output or to change the formulation for a newly emerging virus. The HHS and other organizations worldwide are supporting efforts to develop cell-culture methods for vaccine production. Replacing the traditional method, and providing the capacity for rapid scale-up, could allow vaccine manufacturers to keep pace with evolving flu viruses.

A long-term project at the NIH has been initiated to develop live-attenuated vaccines for each of the known avian flu virus subtypes. This effort, the first of its kind, represents a large-scale, long-term commitment to develop a rapid response capacity to emerging pandemic flu viruses.

Clinical lots will be generated from each reference virus and evaluated in humans for safety, infectivity and immunogenicity. Such an approach will, in principle, allow the rapid production of an avian flu vaccine against any known avian influenza subtype. In the event of a pandemic, vaccine production could theoretically commence almost immediately upon identification of a pre-tested, pandemic seed virus that most closely matched the newly emerged pandemic virus. The timetable for bringing such a vaccine to the market would depend on the pre-existing guidelines for safety testing.

In addition, NIH-supported researchers and others are attempting the difficult challenge of developing a flu vaccine based on antigens that are conserved across several influenza A strains. Such a vaccine could be more broadly protective against whatever strain an individual might encounter⁵.

Prophylactic vaccination with 'prototype' antigens might provide partial protection against a pandemic strain⁶. This concept has been referred to as 'pre-priming', and may offer a way to integrate seasonal flu preparedness with pandemic flu preparedness by adding a fourth strain, such as an H5 or H7 subtype, to the familiar annual trivalent flu vaccine. Of course, these ideas are exploratory and preliminary; studies and discussion are essential to discern whether this approach makes sense from scientific, safety and logistical perspectives.

Fragile enterprise

Strong collaborations between governments, academic institutions, and industry are needed to ensure a reliable vaccine supply. At present, pharmaceutical companies are reluctant to enter or remain in the business of manufacturing vaccines — unpredictable consumer demands and lack of financial incentives make vaccine manufacture a risky business. The biomedical research community can help by developing state-of-the-art technologies. These should be shared with industry to streamline the manufacturing process and make it more flexible, predictable and able to adapt to evolving viruses. Financial and economic incentives (including fair pricing, guaranteed purchase of unsold supplies, tax incentives, a streamlining of the complex regulatory processes needed for vaccine licensure, liability protection), and a vigorous public education effort are under consideration to ensure a steady supply and demand for vaccines.

The evolving situation in Asia offers an unprecedented opportunity to prepare for avian viruses with pandemic potential. Unlike the situation before previous flu pandemics, we now have the knowledge and technology to develop countermeasures for this deadly disease. However, unless we improve our capacity to produce such countermeasures, we may experience again the devastation of past pandemics. ■

Anthony S. Fauci is at the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892-2520, USA.

1. Li, K. S. *et al.* *Nature* **430**, 209–213 (2004).
2. Webby, R. J. *et al.* *Lancet* **363**, 1099–1103 (2004).
3. Choi, Y. K. *et al.* *J. Virol.* **78**, 8609–8614 (2004).
4. Clements, M. L., Betts, R. F. & Murphy, B. R. *Lancet* **1**, 705–708 (1984).
5. Kilbourne, E. D. *Nature Med.* **5**, 1119–1120 (1999).
6. Gensheimer, G. F. *et al.* *Emerg. Infect. Dis.* **9**, 1645–1648 (2003).

Acknowledgement: The author thanks N. Touchette, H. Harvey, G. Folkers, B. Gellin, B. Murphy, K. Subbarao, L. Lambert, M. Giovanni, B. Mulach and S. Foster for helpful discussions.

BOOKS & ARTS

A change of mind?

Putting evolutionary psychology to the test.

Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature

by David J. Buller

Bradford Books: 2005. 552 pp.

\$34.95, £22.95

Oliver Curry

Evolutionary psychology argues that the human mind is a collection of special-purpose circuits designed by natural selection to solve the problems of survival and reproduction that were recurrent in the lives of our ancestors — problems such as finding food, picking habitats, attracting mates and navigating the social world.

In *Adapting Minds*, the philosopher David Buller aims to show that evolutionary psychology is “wrong in almost every detail”. He argues that it is based on a mistaken view of neural development, that its reconstructions of the environment in which humans evolved are “pure guesswork”, and that its major empirical findings are better explained by alternative theories. However, despite this barrage of criticism, Buller’s attempted demolition ultimately fails.

First, Buller relies on the theory of ‘neural darwinism’ to argue that the functional organization of the brain is the product not of genetic instructions, but of a process analogous to natural selection that occurs during the lifetime of an individual. Buller claims that genes merely provide an initial over-supply of neurons and connections — a formless “mass of clay”. These neurons then engage in a darwinian fight to the death, from which “circuits will develop that are specialized in dealing with whatever environmental inputs are most salient”. Thus the mind is not adapted to ancestral conditions; it is capable of adapting to whatever the immediate environment demands.

However, the notion that development resembles a darwinian struggle is spurious. To produce adaptations, natural selection requires numerous iterations of random variation and differential replication among competing entities. There is no equivalent process in development: neurons are not generated at random, ‘successful’ neurons are not reproduced, and the process is not repeated. As a result, there is no prospect of cumulative, adaptive evolution. Indeed, far from being in competition, from a genetic point of view,

neurons are on the same team. Development is better seen as the successful solution to a vast coordination problem involving an elaborate division of labour.

The deeper problem with this account of development is that it supposes that neurons ‘respond’ to the environment in some general way. But, as Buller recognizes, such brains “would face the insoluble problem of learning which of the world’s features are worth learning about before they set about learning about them”. Later, Buller concedes that development must be guided by specific “innate hypotheses” about what to attend and respond to in the world. But the hypotheses that he proposes are woefully inadequate for the task. He suggests, for example, that the ability to recognize faces and facial expressions could begin in infants with nothing more than the hypothesis: “Triangulated high-contrast blobs are very important.” This might direct the child to look in the right direction. But then what? Without further innate guidelines, development would stall for exactly the reason that Buller suggests. The development of each ability must be guided by its own rich battery of innate hypotheses. But with that, we arrive

back at what evolutionary psychology had in mind all along.

So the question is not whether the brain exhibits innate structure, but of what that structure consists. This brings us to the second problem with Buller’s analysis. He is fiercely critical of the methods that evolutionary psychologists use to investigate the selection pressures at work during human evolution. But he undermines these criticisms by using these same methods himself — and to good effect. For example, he argues that male sexual jealousy is a function of squandered “mating effort” rather than “uncertainty of paternity”, and that much of the abuse of stepchildren is best viewed as a form of maternal infanticide. However, in doing so, Buller demonstrates — despite his earlier scepticism — how the careful use of logic, game theory, comparative psychology, primatology and studies of hunter-gatherers can be used to construct plausible scenarios of ancestral conditions, and to generate testable predictions about human psychology and behaviour.

Finally, Buller’s empirical criticisms tend to promise more than they deliver. He states, for example, that there is “no convincing evidence”

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Looking for cues: infants have an innate ability to recognize faces and expressions.

that men have an evolved preference for mating with young women. But what he actually argues is that some men find young women attractive all the time, and that all men find young women attractive some of the time, but that not all men find young women attractive all of the time. This is perfectly consistent with the standard view from evolutionary psychology, in which a preference for youth is only one among many aspects of evolved male sexual psychology.

Adapting Minds is destined to become required reading among evolutionary psychology's detractors. But, despite its flaws, it will be read with interest by evolutionary psychologists too. Buller provides a useful overview

of the field and of the current debates. He challenges evolutionary psychologists to re-examine which of their theoretical commitments are important and why. He advances alternative evolutionary hypotheses, which, far from replacing evolutionary psychology, could contribute to its ongoing refinement. And, above all, by eschewing the personal and political mudslinging that characterized earlier debates over sociobiology, Buller enables evolutionary psychologists to get back to arguing about the science. ■

Oliver Curry is in the Workshop in Political Theory and Policy Analysis, Indiana University, 513 North Park, Bloomington, Indiana 47408-3895, USA.

EXHIBITION

Engineering space-time

Albert Einstein: Ingenieur des Universums/Chief Engineer of the Universe

At Kronprinzenpalais, Berlin until 30 September 2005
www.einsteinausstellung.de

Alison Abbott

"Dr Albert Einstein, Chief Engineer of the Universe, School of Advanced Study, Princeton..." An envelope so addressed is one of more than a thousand items now on display at the Kronprinzenpalais in Berlin.

Did Einstein consider himself American? Or Swiss? Or German? What should we make of this Jew who first renounced German citizenship in 1896, when he was studying in Zurich, who fled from Berlin and the Nazis in

1933, and whose reputation in his homeland was smeared by many in his own scientific community? Is it not perplexing that Germany has put on the world's most ambitious — and arguably the most impressive, thoughtful and imaginative — exhibition commemorating Einstein's life and scientific achievements, implicitly claiming him as their own?

The exhibition is challenging, and in a world used to dumbing down, it may be seen as an élite indulgence. Only visitors with some background in physics who are comfortable with abstract concepts will be rewarded, but they will find it extremely enriching. It presents Einstein and his science in many different contexts: his life and loves in politically turbulent times, for example, but most pertinently, the state of science when he began his working life.

The first of three exhibition floors in the old palace, converted at a cost of €4 million (US\$5 million), is dedicated to setting the scientific scene. A core concept is the history of the development of scientific ideas, elegantly demonstrated in two ground-floor rooms, one dedicated to invisible forces such as magnetism and electricity, the other to the Universe. The rooms display historic instruments, from Galileo's telescope to Otto Hahn's nuclear fission equipment, set among interactive computer terminals that access deeper levels of digital information, including films and sound recordings of some of the scientists. These rooms show how physics concepts developed, one discovery at a time, bringing the visitor to the point at which Einstein entered the fray.

The room near the entrance is empty, its space filled only with one of those impossible debates about which scientists like to fantasize. On three of its walls the projected figures of Aristotle, Newton and Einstein discuss gravity. The actors convey the scientists' personalities, as well as their understanding of the physical world. Introducing themselves, Einstein respectfully hails Aristotle as a true researcher, while Newton arrogantly sneers at Aristotle's "belief in fairy tales".

The second floor is dedicated to Einstein's life, childhood, family, politics, and the social and scientific realities of his world. A second impossible debate takes place, this time between Ludwig Boltzmann, Hendrik Lorentz and Max Planck, frustrated by the apparent dead-end that their physics had reached. The stroke of genius that shifted physics into higher gear — Einstein's special theory of relativity — is celebrated in the central room of the exhibition space. This is lit dazzlingly white and contains only a single central column, an elaborate interactive digital display that illustrates and explains the theory.

The third floor is dedicated to Einstein's legacy, the influence of his work on science and culture today. It is a jamboree of modern experimentation with direct connections to experiments at laboratories such as CERN and the European Southern Observatory.

Nearly halfway through this centenary of Einstein's *annus mirabilis*, and with Einstein-fatigue already setting in, it is still worth spending hours, or even days, at this exhibition. It has an intellectual depth rarely attempted in exhibitions today, and is well served by its elegant presentation. But visitors will have to sweat for their pleasure. They will need the students ('explainers'), present in each room for the exhibition's five-month run, to help them find their way through the maze of information.

Poignantly, the exhibition opened a week after Berlin's Holocaust memorial opened to the public nearby. And many of the artefacts on display were provided by the Hebrew University of Jerusalem, which has a remarkable Einstein collection. Time provides perspective and equilibrium.

Those who pay the exhibition the attention



The ghost of Albert Einstein haunts the room of invisible forces at a Berlin exhibition in his honour.

Bright idea: a white room celebrates Einstein's special theory of relativity.



BERNO BUFF > FOTOGRAFIE

it deserves will note Einstein's opening, gently teasing, gambit to Newton: "I had the honour of standing on your shoulders and thus being able to see a little further". Einstein also stood on the more modern German shoulders of Boltzmann, Planck and Ernst Mach. But Einstein, so the exhibition tells us, belongs to international science — and this shaper of space-time was shaped by his own space and time. ■

Alison Abbott is *Nature's* senior European correspondent.

A braver, newer world

Never Let Me Go

Kazuo Ishiguro

Faber and Faber/Alfred Knopf: 2005.

263 pp. £16.99/\$24

Justine Burley

Never Let Me Go is a literary tour de force and the finest expression of moral disquietude over advances in the biomedical science since Aldous Huxley's *Brave New World* more than 70 years ago. Whereas Huxley's dystopia is central to his story, in Kazuo Ishiguro's tale the dystopic social order provides menace from the background. There are, however, notable likenesses between the two books. In both, the characters are bred and educated to fulfil predestined roles. In both, the changes that have been wrought by science on our moral landscape are traceable to bald utilitarian reasoning. And in both, a misguided notion of progress is laid bare by juxtaposing an existing world against an alternative future one.

Whereas the title of *Brave New World* dangles the lure of progress, with sardonic wit, the title of *Never Let Me Go* beseeches us to cling tightly to our current values. It is no accident that Ishiguro's book is set in the 1990s when the birth of Dolly the cloned sheep was announced and the human genome project was well under way. Both developments prompted debate over what limits

should be placed on applications of related technologies. The prospect of having detailed control over the sort of people who come to exist made many uneasy. For the first time since Darwin's theory of evolution took the world by storm, people sensed that our ethical parameters might be starting to shift. In *Never Let Me Go*, such a shift has occurred: the maxim that individuals should never be treated solely as a means to the ends of others is outmoded.

The few facts we pick up about how this transition occurred, and about the complexion of the resulting society, are revealed gradually, in the telling of a life. The narrator is Kathy, a carer who is soon to be "retired". Her two friends, Tommy and Ruth, have already "completed" following their "donations". Kathy's reminiscences of their intertwined lives return time and again to their cloistered upbringing at a boarding school called Hailsham, unveiling in the process the appalling fate for which it prepared them. The most powerful aspect of the story is the author's intensely controlled psychological portrait of these three individuals, rather than the raw facts about the system that oppressed them.

Much attention is paid to banalities minutely observed: a sports pavilion, an incident involving a lost pencil case, and isolated acts of tomfoolery. These preoccupations of memory strike the reader as being at once utterly normal and utterly incomprehensible. It is as necessary to this book as it was to Ishiguro's *The Remains of the Day* that we have a detailed insight into how his characters see the world and yet never manage to grasp quite how they can see it thus. This imposed distance between reader and subject serves to force a confrontation with big questions about human identity and all that shapes our values, the temptations of science and technology included.

The only coercive force ever named in the book is 'society'. Misuse of technology in the name of the common good is by no means a novel theme. Unique to Ishiguro's treatment of it is his carefully layered picture of the way in which the principal characters forge and retain their individuality in circumstances designed to deny them any importance as individuals. What most chills the reader is not the cold reality they face, it is that they have internalized it and made it a part of who they are.

Never Let Me Go ought to win every major literary award this year. It captures so very beautifully the importance of what we care about. ■

Justine Burley is at the Graduate School for the Integrative Sciences and Engineering, National University of Singapore, MD 11, #01-10, 10 Medical Drive, 117597, Singapore.

MORE ON EINSTEIN

Here is a selection of the new and reissued books about Einstein published this year to celebrate the World Year of Physics.

For those who wish to read the five papers that Einstein produced in 1905, **Einstein's Miraculous Year** edited by John Stachel (Princeton University Press, \$16.95, £10.95) is a good place to start. This new version contains a foreword by Roger Penrose and a new introduction by John Stachel examining Einstein's early life, leading up to the writing of these famous papers.

Alice Calaprice, author of **The Quotable Einstein**, has now produced **The Einstein Almanac** (Johns Hopkins University Press, \$24.95, £17), a chronological listing of 300 of Einstein's publications spanning the years from 1901 to 1955. These are interspersed with biographic details and other notable news from the world of physics. This is a useful book for Einstein enthusiasts to dip into. Examples of Einstein's more popular writings can be found in the reissued **Ideas and Opinions** (Souvenir Press, £9.99).

For those seeking to understand the significance of Einstein's papers, Nigel Calder's **Einstein's Universe** (Penguin, £7.99, \$14) has been reissued. **Einstein's Cosmos** by Michio Kaku (W. W. Norton, \$13.95) examines how Einstein's vision transformed our understanding of space and time. A second edition of David C. Cassidy's **Einstein and Our World** (Humanity Books, \$21) places Einstein's work in its scientific and wider cultural context.

Einstein: A Life in Science by Michael White and John Gribbin (Free Press, £7.99), a short biography, makes a reappearance, along with another offering by John Gribbin, this time with Mary Gribbin, **Annus Mirabilis** (Penguin/Chamberlain Bros, \$25.95). This brief biography also includes a reproduction of Einstein's own book **Relativity: The Special and General Theory**, first published in 1916, and a DVD of the A&E biography of Einstein.

Reviews of other Einstein books appeared in the 20 January issue of *Nature* (433, 195–197; 2005).

Beating heart disease

A Change of Heart: How the People of Framingham, Massachusetts, Helped Unravel the Mysteries of Cardiovascular Disease

by Daniel Levy & Susan Brink

Alfred Knopf: 2005. 247 pp. \$26.95

Peter Sleight

The Framingham Heart Study began in 1948 and has run for nearly 60 years. By documenting the health and lifestyles of a small New England community, it identified many of the risk factors, such as blood pressure, cholesterol and smoking that are now associated with heart disease and stroke. In *Change of Heart*, Daniel Levy, the current director of the Framingham study, and journalist Susan Brink tell the history of this epidemiological research. The combination of history, anecdotes and the political intrigue necessary for the Framingham study to survive should interest professionals as well as the lay public at which this book is aimed.

This book holds many lessons for the present, as we're now seeing huge increases in funding for basic research — but at the expense of clinical and epidemiological studies. Distinguished scientists predict a future in which basic research will be rapidly translated into therapeutic trials, and then to clinical practice, as exemplified by the statins. These academics now dominate the committees that distribute research money, and they naturally favour basic research. They give scant credit to those who identified their laboratory targets — clinicians and epidemiologists such as Ancel Keys, whose 'seven countries' study linked cardiovascular disease to the consumption of saturated fats. Huge amounts of money are spent on genomic research.

Levy and Brink speculate that known risk factors explain only half the causes of heart attack, and that genomic research might identify new targets. But the recent INTERHEART case-control study of heart attack in 52 countries, run on a shoestring budget by Salim Yusuf of McMaster University in Canada and colleagues, shows that 90% were predicted by known risk factors, many of them documented in the Framingham database.

Change of Heart sympathetically describes the naive enthusiasm and ideals of the early Framingham pioneers. Roy Dawber and Bill Kannel thought that giving 5,000 middle-aged people from a 'typical' US community 20 annual medical examinations would reveal the causes of the increasing scourge of heart attacks. They later took advantage of emerging technology, such as electrocardiograms and ultrasound, to add to their clinical and blood data. They also soon realized that they needed a crash course in data handling and epidemiological methods. In the early years, the data



Inching forward: the Framingham study measured body fat in the search for heart-disease risk factors.

were all on paper and were legally controlled by the National Heart, Lung and Blood Institute (NHLBI), which did not always cooperate by releasing it to the Framingham staff. Fortunately, Dawber and Patricia McNamara kept duplicate copies. The project then used IBM punch cards, and finally modern information technology.

Another lesson for our times was the somewhat arbitrary administrative decision by the NHLBI to axe the study in 1968, just at the time when the endpoints — deaths, heart attacks and strokes among those enrolled in the study — were beginning to increase. The study was saved by the subsequent outcry from both subjects and researchers, including cardiologist Paul Dudley White, who wrote to President Nixon.

The four directors of Framingham, Dawber, Kannel, William Castelli and Levy, have preserved the original ethos of putting the subjects first. The subjects developed their own representative board and, more important, gave their time and data for the benefit of medical research. They managed to derail a plan by Framingham Genomic Medicine, a collaboration between Boston University and a group of rich venture-capitalists, to market the data and DNA samples to interested researchers and drug companies. But the collaboration was faced with the threat of the withdrawal of many participants, who first read of the planned deal in the newspapers, with no prior consultation with either project staff or subjects. Boston University finally withdrew from the project. Framingham survived the crisis, and the subjects agreed that their anonymized data and DNA should be a freely available database for scientific research.

Inevitably, there are some weaknesses in the Framingham study. Key among these is the relatively small number of subjects — commentators have remarked that there seem to be more publications on Framingham than there are subjects in the study! Another problem is that the data from this New England community are not universally applicable to other areas or countries. Country- and community-specific risk data have now been obtained elsewhere, by the European SCORE project, for example.

On the other hand, the strengths of the study are immense. No other study has data from three generations of subjects, for instance. The choice of the community of Framingham (near Boston) was fortunate, too — unlike other possible choices in 1948, it had the support of a local university, which proved crucial in the 1968 closure crisis. And it has been very well managed, has fostered great loyalty, and few subjects have been lost from follow-up studies.

The Framingham study had some spectacular successes, notably the continuous relationships between blood pressure, cholesterol, cigarette consumption and long-term risk. This led to landmark hypertension trials carried out by Ed Freis, which were delayed for years by dogmatic assertions that treatment would be dangerous or that a trial would be unethical.

All in all, *A Change of Heart* is an easy but exciting read. We owe a lot to Framingham, and the staff and subjects deserve our gratitude. The book will be a bestseller in Boston and deserves to be so elsewhere. ■

Peter Sleight is in the Department of Cardiovascular Medicine, University of Oxford, Oxford OX3 9DU, UK.

The great chain of being

Our persistence in placing ourselves at the top of the Great Chain of Being suggests we have some deep psychological need to see ourselves as the culmination of creation.

Sean Nee

For centuries the 'great chain of being' held a central place in Western thought. This view saw the Universe as ordered in a linear sequence starting from the inanimate world of rocks. Plants came next, then animals, men, angels and, finally, God. It was very detailed with, for example, a ranking of human races; humans themselves ranked above apes above reptiles above amphibians above fish. This view even predicted a world of invisible life in between the inanimate and the visible, living world, long before Antonie van Leeuwenhoek's discoveries. Although advocates of evolution may have stripped it of its supernatural summit, this view is with us still.

Common presentations of evolution mirror the great chain by viewing the process as progressive. For example, in their book *The Major Transitions in Evolution*, John Maynard Smith and Eors Szathmáry take us from the origin of life, through to the origin of eukaryotic cells, multicellularity, human societies and, finally, of language. They explicitly point out that evolution does not necessarily lead to progress, and even refer to the great chain by its Latin name, *scala naturae*. But it is impossible to overlook the fact that the 'major' evolutionary transitions lead inexorably, step by step, to us. Similarly, in their recent essay in *Nature*, 'Climbing the co-evolution ladder' (431, 913; 2004), Lenton and colleagues illustrate their summary of life-environment interactions through the ages with a ladder whose rungs progress through microbes, plants, and, at the top, large animals.

In his recent book *The Ancestor's Tale*, Richard Dawkins reverses the usual temporal perspective and looks progressively further back in time to find our ancestors. Like Maynard Smith and Szathmáry, he cautions us against thinking that evolution is progressive, culminating with us. He emphasizes that with whatever organism we begin the pilgrimage back through time, we all are reunited at the origin of life. But by beginning the journey with us and looking backwards along our ancestry, Dawkins generates a sequence of chapter titles that would read like a typical chain to a

medieval theologian, albeit with some novelties and the startling omission of God.

By starting with us, Dawkins regenerates the chain because species that are more closely related to us are more similar as well, and such similarity was an important criterion in determining the rankings in the classical chain. But there is nothing about the world that compels us to think

Changes in ocean ecosystems wrought by Bacteria and Archaea contributed to the deposition of the ocean sediments, an event of enormous significance: these sediments became the habitat for bacteria that now constitute about one-third of the total living biomass today. (A side-effect of the deposition is the oxygenation of the atmosphere by photosynthetic bacteria.)

Evolution continued for billions of years, with many remarkable innovations stimulated by both cooperation and conflict. For example, Bacteria evolved the capacity to communicate chemically to coordinate attacks on others, and a willingness to commit suicide for the greater good of the community. Around a billion years ago, a great experiment occurred: Bacteria and Archaea came together in a fusion event to synthesize a whole new domain of life, the Eukarya. Sadly, the outcome was rather uninteresting: the resulting organisms displayed a very limited metabolic repertoire and much restricted habitat requirements.

Over the past 600 million years the Bacteria, Archaea and microbial Eukarya have continued to evolve into brand new niches. As it happens, a few branches of Eukarya — plants and animals — grew freakishly huge bodies. They also created both new substances for bacteria to exploit, such as plant lignins, and new environments for microbes to inhabit, such as feathers and urinary tracts. Indeed, some of the richest and most interesting ecologies on Earth can be found inside the animal gut.

One of the huge species, *Homo sapiens*, got remarkably self-important. But when, to his surprise, a virus wiped him out, most of life on Earth took no notice at all. ■

Sean Nee is at the Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK.

FURTHER READING

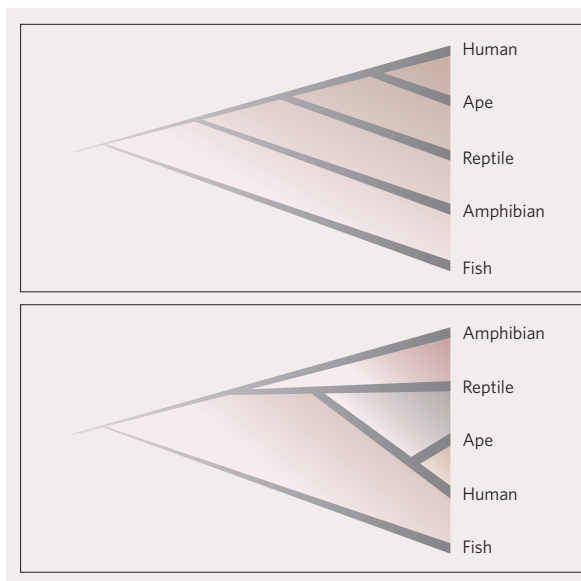
Lovejoy, A. O. *The Great Chain of Being* (Harper and Row, New York, 1965).

Gee, H. *Nature* **420**, 611 (2002).

Maynard Smith, J. & Szathmáry, E. *The Major Transitions of Evolution* (W. H. Freeman & Co., Oxford, 1995).

Dawkins, R. *The Ancestor's Tale* (Weidenfeld & Nicolson, New York, 2004).

Nee, S. *Nature* **429**, 804–805 (2004).



Although both representations are equally valid, we instinctively position ourselves at the top of phylogenetic trees (upper panel).

about it in this way, suggesting, instead, that we have some deep psychological need to see ourselves as the culmination of creation. Illustrating this, when we represent the relationships between species, including ourselves, in a family tree, we automatically construct it so that the column of species' names forms a chain with us as the top, as in the first of the trees pictured. But the other construction is equally valid.

Here is another view of evolution, but this time from the point of view of microbes — the main form of life on our planet. From the mists of time, nearly 4 billion years ago, three great domains of life emerged: Bacteria, Archaea, and the molecular parasites of these, such as viruses. Over hundreds of millions of years the Bacteria evolved an extraordinary variety of biochemical capabilities, including the ability to generate light, and to 'eat' and 'breathe' metals. The Archaea also evolved remarkable capacities to thrive in every environment available, including superheated, pressurized water deep in the oceans.

NEWS & VIEWS

A. HIBBERT/ECOSCENE/CORBIS



Figure 1 | Transport connection. Gilbert *et al.*² find that the geographical distribution of bovine tuberculosis is associated with the number of cattle imported from infected areas.

EPIDEMIOLOGY

Dangers of moving cows

Mark E. J. Woolhouse

The movement of cattle around the country, and the presence of badgers, are both implicated in the high incidence of bovine tuberculosis in Britain. The problem may get even worse in the near future.

Every year in Britain, as in many other countries, millions of cattle are moved between livestock farms, markets and abattoirs. This practice is known to contribute to the spread of infectious diseases¹, and a study by Gilbert *et al.* (page 491 of this issue)² suggests that it is implicated specifically in recent increases in the incidence of bovine tuberculosis (BTB).

Gilbert and colleagues provide the most comprehensive analysis yet of the role of cattle movements (Fig. 1) in the epidemiology of BTB. Their work is a tour de force in the integrated statistical analysis of a set of complex databases, including several million records from the recently created British Cattle Movement Service. This is one of the first times these data have been used for epidemiological research, reflecting the increasing willingness of UK government services to share such resources with the scientific community.

Using a statistical technique known as multiple logistic regression, Gilbert and colleagues showed that the geographical distribution of BTB was most closely associated with the number of recent cattle imports from infected areas. They then used the results from their statistical analysis as inputs into a computer model (in technical jargon, a spatially explicit, stochastic simulation model) of the changing distribution of BTB in Britain. The model predicted the presence or absence of BTB in any given 5-km × 5-km cell in 2003 with more than 80% accuracy, and was similarly successful in predicting the spread of BTB to new cells. Looking further ahead to 2005, they predicted a continued increase in the number

of cases annually (Fig. 2, overleaf), and a risk of further spread from the disease's current strongholds in the southwest and parts of central England to regions such as Wales, Cumbria and the Scottish Borders.

Unsurprisingly, cattle movements were not the only determinant of BTB incidence. Other predictors identified (from a set of 100 possibilities investigated) included the recent presence of BTB locally and measures of climate, land use and vegetation. In a further analysis, Gilbert *et al.* found that movements played a more obvious role outside 'core' areas where BTB is already established, implying that movements are more important for the spread of infection than for its persistence. Indeed, there are regions where BTB occurs only sporadically despite regular imports of cattle from infected areas. Some other necessary factor seems to be missing — which brings the discussion round to wildlife reservoirs, and especially the badger.

Badgers are known to carry BTB, but Gilbert and colleagues' study is equivocal on the animals' role in the epidemiology of the disease in British cattle. Proximity to known badger locations does appear as a predictor in their analyses, but it is not prominent. Interpreting this result is difficult because the available badger data are of patchy quality and the presence of badgers is likely to be correlated with other factors. To better understand the role of badgers in the persistence of BTB, we need to turn to experimental studies, such as the 'Four Counties' trial in Ireland, the results from which were published earlier this year³.

This study compared the rate of detection of BTB in cattle herds in each of four pairings of a reference area (where few badgers were culled) and a removal area (where badgers were culled proactively, regardless of infection in the cattle). By the end of the study period the chances of a confirmed BTB case in a herd in the removal areas had fallen by between 62% and 95% relative to the reference areas. This is the clearest demonstration to date of a link between badgers and BTB in cattle.

It is worth noting that the success of the Four Counties trial was not immediate. Over the five-year study period, in every removal area the number of herds affected sometimes went up, not down, from one year to the next, although a reduction was always achieved eventually. This is no surprise: such variability had already been predicted by mathematical models of the impact of badger culling⁴.

All this provides pointers for a major ongoing experimental study in southwest England, the Randomised Badger Culling Trial. This trial originally had three arms: no badger culling, proactive culling and reactive culling (undertaken only after confirmation of BTB in local cattle). The reactive-culling arm was abandoned because of a reported 27% increase in herds with BTB (relative to the no-culling arm) rather than the anticipated decrease⁵. However, the increase was not statistically significant, indicating that it could simply have been a blip, and that it was arguably still too soon in the course of the trial to expect much of an effect anyway⁶. So the aborted arm tells us little. At least the proactive arm of the trial

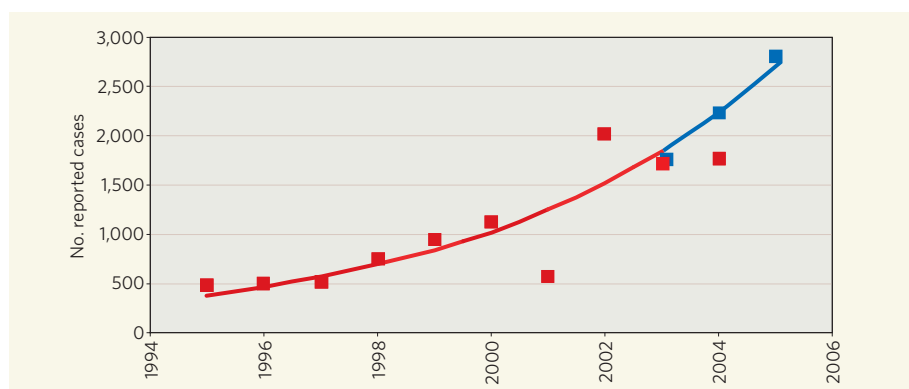


Figure 2 | Trends in the numbers of reported cases of bovine tuberculosis in British cattle 1995–2005. Red squares, observed values, including provisional 2004 data accessed on 8 April 2005. The dip in 2001 is probably due to decreased testing during the epidemic of foot-and-mouth disease at that time. Red line, exponential trend up to 2003. Blue line, extrapolation of exponential trend to 2005; blue squares, average predicted values for 2003–05 from the model of Gilbert *et al.*².

survives, but definitive results are not expected for some years.

The experiences of these trials illustrate an epidemiological principle. The dynamics of chronic infections — such as BTB and BSE in cattle, scrapie in sheep, or tuberculosis and HIV/AIDS in humans — are inherently slow, and it may be years, or even decades, before the effects of any interventions (or any other changes in their epidemiologies) are fully realized. We should not expect a quick fix.

That said, given that the incidence of BTB in Britain is expected to remain high or even increase further (Fig. 2), the sooner that more effective control measures are introduced, the better. The UK government is reviewing its options for tackling BTB⁷; these include both badger culling and the statutory pre- and post-movement testing of cattle (which

would benefit from improved diagnostics), and might conceivably extend to greater restrictions on cattle movements. The research discussed here suggests that all of these approaches merit serious consideration. ■

Mark E. J. Woolhouse is at the Centre for Infectious Diseases, University of Edinburgh, Easter Bush, Roslin, Midlothian EH25 9RG, UK. e-mail: mark.woolhouse@ed.ac.uk

1. Woolhouse, M. E. J. *et al. Biol. Lett.* doi:10.1098/rsbl.2005.0331 (in the press).
2. Gilbert, M. *et al. Nature* **435**, 491–496 (2005).
3. Griffin, J. M. *et al. Prev. Vet. Med.* **67**, 237–266 (2005).
4. Smith, G. C. *et al. J. Appl. Ecol.* **38**, 509–519 (2001).
5. Donnelly, C. A. *et al. Nature* **426**, 834–837 (2003).
6. Roper, T. J. *Nature* **426**, 782–783 (2003).
7. DEFRA Government Strategic Framework for the Sustainable Control of Bovine Tuberculosis (bTB) in Great Britain (Dept for Environment, Food & Rural Affairs, London, 2005).

PLANETARY SCIENCE

When giants roamed

Joe Hahn

An early epoch of planetary migration could explain the current orbits of the giant planets, the origin of Jupiter's Trojans, and an intense bombardment of the early Solar System with a shower of asteroids and comets.

In a triplet of papers in this issue^{1–3}, H. F. Levison and colleagues contend that the orbits of Jupiter, Saturn, Uranus and Neptune have been disturbed in a small but significant way. They argue that the giant planets' eccentricities (the deviation of their orbits from a true circle) and inclinations (the tilt of their orbital planes) are much larger than those predicted by theories of planet formation — which implies that some process has disturbed the orbits of the giant planets since the time of their formation.

In the first paper (Tsiganis *et al.*, page 459)¹, the authors show that the passage of Jupiter and Saturn through a 1:2 mean-motion resonance

(MMR) can account for the orbital spacings, eccentricities and inclinations of all four giant planets. An MMR is a 'sweet spot' in the Solar System when the orbital periods of two bodies are ratios of whole numbers, and this is also a site where a planet's periodic gravitational perturbations can 'pump-up' another body's eccentricity and/or inclination. Thus Jupiter's 1:2 MMR is the site where Saturn completes one orbit about the Sun for every two orbits of Jupiter. The authors' find that the passage of Jupiter and Saturn through this resonance can excite their eccentricities and inclinations to current levels. However, Jupiter and Saturn are

currently rather far from the 1:2 resonance — the ratio of their current orbital periods is near 1:2.5 — so the implication here is that these planets have since migrated through 2:1 to their present positions. This is a remarkable concept, because we usually think of the planets' orbits as being rather static and changing little over time.

There is, however, good reason to believe that the giant planets' orbits did undergo a marked readjustment early in the Solar System's history. The best evidence for planetary migration is preserved in the Kuiper belt, which is a swarm of comets orbiting beyond Neptune. Several of these objects are also in eccentric orbits at Neptune's 3:2 MMR, where they circle the Sun twice for every three orbits of Neptune. However, these bodies are unlikely to have formed in such unusual orbits; rather, the prevailing thinking is that they were gravitationally trapped at this resonance while Neptune migrated outwards⁴ early in the Solar System's history.

But what could have caused the planets' orbits to migrate? During this early epoch of planet formation, interplanetary space was still filled with numerous bits of debris — planetesimals — that had not yet been accreted by the newly formed planets. But as the giant planets grew to their final sizes, their great masses made them very effective at tossing this planetesimal debris around the Solar System, an interaction that also caused the planets' orbits to shift in response. In the model of Tsiganis *et al.*, Neptune's orbit doubles in size when it is gravitationally scattered by Saturn into a more distant and eccentric orbit about the Sun (Fig. 1). Neptune's subsequent gravitational interactions with the numerous planetesimals then makes its orbit more circular as that planet migrates farther outwards. However, this orbital evolution requires a lot of debris — a mass equivalent to about two Neptunes, all of which must ultimately be scattered out of the Solar System by the migrating planets. Evidently, planet formation was a messy and inefficient business.

The scheme sketched by Tsiganis *et al.*¹ is plausible in the sense that their model does indeed excite the eccentricities and inclinations of the giant planets to the requisite values. But caution is required: the fact that a simulation of planet formation produces an end state in good agreement with the observed Solar System does not prove that the simulated events actually happened. Rather, the eccentricities and inclinations of the giant planets could instead be relics of another process — perhaps arising from gravitational perturbations from other large protoplanets that might also once have roamed the early Solar System. Indeed, Uranus's very large obliquity (its rotational axis is tipped 98° away from its orbital axis) is usually interpreted as evidence for a collision with another protoplanet that had a mass comparable to that of Earth⁵. Nonetheless, the authors' model deserves credit for at least having

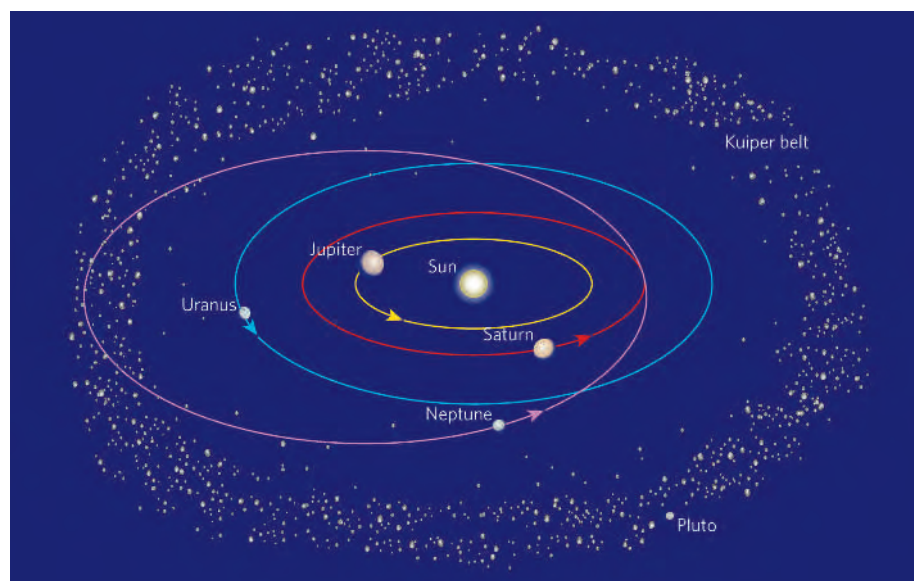


Figure 1 | Cosmic billiards. According to the model of Tsiganis *et al.*¹, Saturn scattered Neptune outwards beyond Uranus and into the Kuiper belt during the epoch of planet migration. The ellipses indicate the orbits of the four giant planets, Jupiter, Saturn, Uranus and Neptune, viewed obliquely as they travel about the Sun, with Pluto lurking out in the Kuiper belt.

been rigorously tested, unlike other proposals.

Another interesting finding is described in the second paper (Morbiddelli *et al.*, page 462)², which shows that this planet migration scheme can also account for the existence of Jupiter's Trojan asteroids. The jovian Trojans inhabit two small, stable niches near Jupiter's orbit, one 60° in advance of the planet, the other trailing by 60° — and these bodies place rigorous constraints on models of planet formation. Indeed, their existence would seem to jeopardize the model of Tsiganis *et al.*, which shows that the passage of Jupiter and Saturn through the 1:2 MMR also relieves Jupiter of all its Trojans. But the model also shows that such flows can go in both directions, because a tiny fraction of the numerous planetesimals (those that are scattering off the planets and thus driving their migration) can still settle back into stable orbits at the Trojan sites, replacing those Trojans that were lost.

The third paper (Gomes *et al.*, page 466)³ shows that the 1:2 MMR can also be implicated in the Late Heavy Bombardment, which was a brief but intense series of impacts known to have occurred early in the Moon's history. In the authors' models, Neptune's orbit is destabilized when Jupiter and Saturn pass through the 1:2 MMR. The subsequent scattering of Neptune by Saturn then tosses Neptune out into the more distant planetesimal disk that surrounds the entire Solar System. Neptune's vigorous gravitational scattering of those planetesimals then sends a burst of impactors throughout the entire Solar System — including some that hit the Moon.

Such a sudden readjustment of the Solar System's architecture could certainly have resulted in a heavy bombardment of all planetary surfaces; the only contentious issue here is in the timing of this event. Laboratory studies of the

lunar rock samples that were collected by the Apollo astronauts show that these rocks were melted by near-simultaneous impacts about 600 million years after the Moon formed⁶. This is seemingly in conflict with other models of Solar System formation, which show Neptune migrating during the first 10 million years after the Moon's formation^{7,8}.

But Tsiganis *et al.* suggest that this migration was delayed by the presence of the solar nebula, the massive disk of gas around the Sun within which the planets formed. Until the nebula had dissipated, they argue, the giant planets' gravitational perturbations would have rid the solar nebula of all planetesimal debris that was orbiting between the planets, yet would still have left a dense belt of planetesimals orbiting just beyond the outermost giant planet. And because those more distant planetesimals were orbiting just at the edge of Uranus and Neptune's gravitational 'grasp',

their gravitational interactions were greatly reduced and thus the planetary migration proceeded much more slowly — which accounts for the delay in the 1:2 passage that ultimately triggered their late Late Heavy Bombardment.

The success of the model of Gomes *et al.*³ hinges on the giant planets' ability to clear the solar nebula of its interplanetary debris before the nebula itself has dispersed. However, a planet's gravity does not discriminate between a parcel of gas and a solid planetesimal, so it is unclear whether this could, in fact, have happened. Indeed, other models show that a giant planet tends to launch spiral density waves into the nebula gas⁹, and this wave-action tends to smear the planets' perturbations across large parts of the gas disk. So it is possible that the planetesimals might instead have persisted by simply bobbing about as these waves washed over them.

Even so, the work of Gomes *et al.* does raise an interesting hydrodynamic problem that deserves closer scrutiny. If further studies show that the newly formed giant planets could indeed have depleted the solar nebula of its residual solids, the authors will have arrived at a compelling explanation for the Late Heavy Bombardment of the Moon — as well as for the structure of the outer Solar System. ■

Joe Hahn is at the Institute for Computational Astrophysics, Saint Mary's University, Halifax, Nova Scotia B3H 3C3, Canada.
e-mail: jhahn@ap.smu.ca

1. Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. *Nature* **435**, 459–461 (2005).
2. Morbidelli, A., Levison, H. F., Tsiganis, K. & Gomes, R. *Nature* **435**, 462–465 (2005).
3. Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. *Nature* **435**, 466–469 (2005).
4. Malhotra, R. *Astron. J.* **110**, 420–429 (1995).
5. Korycansky, D. G., Bodenheimer, P., Cassen, P. & Pollack, J. B. *Icarus* **84**, 528–541 (1990).
6. Ryder, G. *Eos* **71**, 322–323 (1990).
7. Hahn, J. M. & Malhotra, R. *Astron. J.* **117**, 3041–3053 (1999).
8. Gomes, R. S., Morbidelli, A. & Levison, H. F. *Icarus* **170**, 492–507 (2004).
9. Lin, D. N. C. & Papaloizou, J. C. B. in *Protostars and Planets III* (eds Levy, E. & Lunine, J. I.) 749–835 (Univ. Arizona Press, Tucson, 1993).

DEVELOPMENTAL BIOLOGY

A blank canvas no more

Yoshiki Sasai

Embryonic cells learn their fate early in development. Discovery of a factor that controls the development of one embryonic tissue, the ectoderm, highlights a mechanism that might also influence the growth of cancer cells.

The enormous diversity of body structures in the animal kingdom is somewhat surprising given the commonality of certain developmental processes. The early stages of embryogenesis in bilaterally symmetrical animals, for example, generally involves the process of

gastrulation, when the embryo reorganizes from a simple ball of cells into a multilayered organism with a recognizable body plan. This mysterious process generates the first visibly distinct tissues, the three 'germ' layers of ectoderm, mesoderm and endoderm¹. Dupont

and colleagues, writing in *Cell*², identify a protein factor that determines which cells become ectoderm. The factor also turns out to have a fundamental role in allowing adult cells to proliferate, and there are hints that it may therefore be involved in some cancers.

According to Lewis Wolpert, "It is not birth, marriage, or death, but gastrulation, which is truly the most important time in your life". Indeed, the tissues produced by gastrulation form the basis for all adult tissues except germ cells (eggs, sperm and their precursors). The ectoderm gives rise to the epidermis and nervous systems, the endoderm contributes mainly to the digestive system and its related organs, and the mesoderm develops into the variety of tissues that fill the space between the epidermis and the digestive tract.

In amphibians, the three germ layers form at the beginning of gastrulation, with the ectoderm at the top of the embryo (the 'animal' pole) and the endoderm at the bottom (the 'vegetal' pole) (Fig. 1). As the animal pole of amphibian embryos is pigmented and the vegetal pole is not, the process is easy to follow. Together with the comparatively large size of amphibian embryos at this stage, this has made them a favourite model for embryologists studying gastrulation³.

Studies using the African clawed frog, *Xenopus laevis*, have revealed that two basic molecular mechanisms are at work in the induction of the germ layers: local enrichment of maternally supplied intracellular factors that promote mesodermal and endodermal determination (such as the gene-regulatory factor VegT)⁴, and cellular interactions involving extracellular inductive signals (maternal and embryonic), such as members of the transforming growth factor β (TGF- β) protein

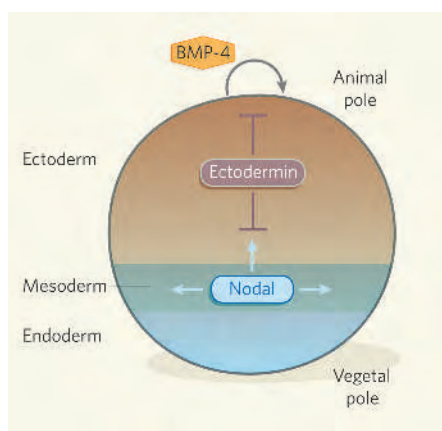


Figure 1 | Gastrulation. In the *Xenopus* embryo, the endoderm, mesoderm and ectoderm develop sequentially from the animal (pigmented) pole, and move towards the vegetal (non-pigmented) pole. The final fate of the cells is largely determined by members of the transforming growth factor β family of proteins, such as Nodal and bone morphogenetic protein 4 (BMP-4). Ectoderm shields the region that is fated to become ectoderm from the influence of these factors, which diffuse freely among the cells.

family⁵. Both mechanisms activate mesodermal and endodermal development, but there are no known extracellular factors that specifically induce the development of ectoderm. This has led to the conventional wisdom that the ectoderm is merely the manifestation of a default state in the absence of VegT and/or inductive signals to the contrary.

But is this really the case? It is a question that occurs to me whenever I detect mesoderm-inducing factors in tissues adjacent to, or even within, the future ectoderm region of a *Xenopus* embryo. For instance, there is a remarkably high level of bone morphogenetic protein 4 (BMP-4), a TGF- β protein that has mesoderm-inducing activity, in the entire animal region (the future ectoderm) of the pre-gastrula embryo⁶.

Dupont *et al.*² now provide clear evidence that ectoderm is not simply the default tissue. A search for *Xenopus* genes that promote ectodermal rather than mesodermal differentiation led to the isolation of a region-specific maternal factor, which the authors have named Ectoderm. This protein is found in cell nuclei, and is required to suppress aberrant mesodermal differentiation in the ectoderm (Fig. 1). In embryos that are depleted of Ectoderm, mesodermal and endodermal inducers appear in the region that would otherwise become the ectoderm.

Ectoderm attenuates the cellular response to mesoderm-inducing signals from the TGF- β proteins BMP-4 and Nodal. It does this by promoting the degradation of Smad4, a protein that directly regulates gene activity in response to TGF- β signals (Fig. 2). In effect, the function of Ectoderm is to ensure that any instructions from wayward mesodermal inducers fall on deaf ears. So ectoderm formation is not simply the realization of a passive, ground state, but is a process that involves at least one active mechanism to neutralize inappropriate developmental cues.

In *Xenopus*, the ectodermal specification mechanism appears to depend on maternally supplied positional information that directs the localization of Ectoderm messenger RNA towards the animal pole of the embryo. Will this mechanism also apply to early mammalian embryos, where such information seems to be in much shorter supply? This is a challenging question. In mice, Nodal has an essential role in mesodermal induction from the epiblast (the layer of uncommitted cells that generates the three germ layers)⁷, but little is known about how mesoderm-inducing signals act selectively on a limited region of the mouse epiblast. A similar mechanism of 'active denial' may also occur in mice, because the early mouse embryo at the stage of germ-layer specification is much smaller than that of the frog, so spatial separation alone is unlikely to be sufficient to insulate the ectoderm from the influence of factors promoting a mesodermal fate.

The TGF- β signalling pathway is notorious

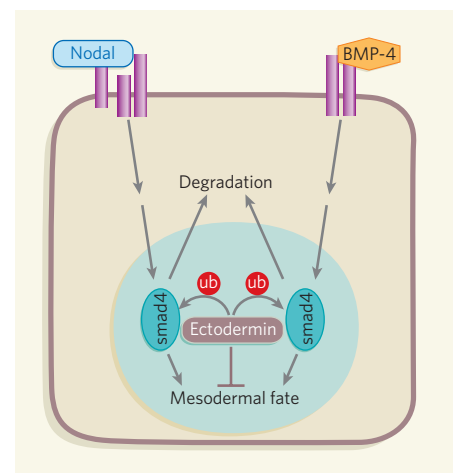


Figure 2 | Denying fate. Nodal and bone morphogenetic protein 4 (BMP-4), members of the transforming growth factor β family of proteins that control cell fate, interact with receptors on the cell surface. This initiates a cascade of signals (particularly phosphorylation of Smad1 and Smad2) that lead to the activation of Smad4, a gene regulator, causing the cell to adopt a mesodermal fate. Ectoderm interferes with this pathway by tagging Smad4 with a ubiquitin molecule (ub), which marks the protein out for degradation. In the absence of the signals from Nodal and BMP-4, the cell becomes ectodermal.

in cancer biology, and Dupont *et al.* find that Ectoderm has a causal role in the proliferation of cancerous cells². The aberrant inactivation⁸ of Smad4 is associated with several types of human malignancy. Interestingly, Ectoderm is expressed abundantly in C32 colorectal cancer cells, which have no mutation in the *Smad4* gene. These cells proliferate because the TGF- β signals that would normally stop them from dividing are somehow disrupted⁸. Depletion of Ectoderm showed that it is responsible for interfering with TGF- β signalling and hence for the burgeoning of the cells.

Given that Ectoderm is a key player in such a basic cell function as proliferation, it could well be meaningless to talk about the presence or absence of Ectoderm activity with regard to the ground state of early embryonic cells. Instead, I would say that even by this stage, each germ layer has a 'colour' of its own. None is simply a blank canvas on which colours are painted.

Yoshiki Sasai is in the Organogenesis and Neurogenesis Group, RIKEN Center for Developmental Biology, 2-2-3 Manatogijima-Minamimachi, Chuo, Kobe 650-0047, Japan. e-mail: sasaicdb@mub.biglobe.ne.jp

1. Wolpert, L. *Dev. Suppl.* 7-13 (1992).
2. Dupont, S. *et al. Cell* **121**, 87-99 (2005).
3. Gerhart, J. *Int. J. Dev. Biol.* **43**, 605-613 (1999).
4. Zhang, J. *et al. Cell* **94**, 515-524 (1998).
5. Agius, E. *et al. Development* **127**, 1173-1183 (2000).
6. Fainsod, A., Steinbeisser, H. & De Robertis, E. M. *EMBO J.* **13**, 5015-5025 (1994).
7. Collignon, J., Varlet, I. & Robertson, E. J. *Nature* **381**, 155-158 (1996).
8. Siegel, P. M. & Massagué, J. *Nature Rev. Cancer* **3**, 807-821 (2003).

ULTRAFAST SCIENCE

Molecular structure in an instant

Jonathan P. Marangos

The observation that there is interference between a laser-induced electron wave and a single molecule means that it may be possible to image changes in molecular structure with a sub-femtosecond resolution.

Scientists have long dreamt of being able to instantaneously image molecular structure with a temporal resolution less than a femtosecond (10^{-15} s) — the timescale of fundamental physical and chemical changes. A technique with the potential to do this is reported by Kanai *et al.*¹ on page 470 of this issue. The measurement uses an intense laser field to tear an electron from a molecule (ionization), accelerate it and then drive it back towards the molecule a fraction of an optical cycle later (about 1 femtosecond). The electron recombines with the molecule, giving up the energy it acquired (typically 30–70 eV) as a single, soft X-ray photon in a process known as high-order harmonic generation.

The returning electron can be considered as a superposition of electron waves of different energy and momentum. The de Broglie wavelength of an electron wave is given by $\lambda = h/p$ where h is Planck's constant and p is the momentum of the electron. The intensity of emission at a particular X-ray photon energy is sensitive to the wavelength of the electron wave and the shape of the wavefunction of the electrons in the molecule. It is possible to use this phenomenon to probe structures because the electron wavelength is comparable to the size of the molecular wavefunction.

When the electron wavelength matches the distance between the nuclei of the molecule, the internuclear separation, interference occurs between the emission amplitudes from different parts of the molecule and modulates the total intensity of soft X-ray emission

(Fig. 1). The value of the photon energy at which the modulation is observed indicates the internuclear spacing at that instant. This type of interference in the high-order harmonic process from molecules was first predicted several years ago², but had not yet been seen.

Kanai and co-workers¹ now report that they have observed this phenomenon. Their experiment involved the measurement of high-order harmonic generation from a gas of linear molecules (N_2 , O_2 or CO_2) with their axes aligned in space. An initial, ultrafast laser pulse, which excites a 'rotational wavepacket' (a coherent superposition of states) in the molecules, controls the molecular alignment³. The rotational wavepacket then exhibits strong molecular-axis alignment at regular periods (every 5.4 picoseconds in the case of CO_2). A second, higher-intensity, ultrafast laser pulse (about 2×10^{14} W cm⁻²) then produces high-harmonic emission from molecules close to the maximum degree of alignment. By precisely varying the delay before the second pulse, the emission intensity for different magnitudes and angles of alignment can be measured.

The set-up and the principle are similar to those employed by Itatani *et al.*⁴, who used a tomographic retrieval technique to image the electronic states of a nitrogen molecule. The advance of Kanai *et al.* is the simultaneous observation of the ionization and the intensity of the soft X-ray emission, allowing the efficiency of the first, ionization step of the process to be separated from the efficiency of

the final step when the electron wave recombines with the molecular electronic state.

The authors found a minimum in the harmonic emission from CO_2 at maximum ionization. In contrast, O_2 and N_2 had maximum harmonic emission at maximum ionization. For CO_2 , the greatly reduced efficiency of the recombination step results from destructive interference. The parts of the molecular electronic state of CO_2 located near the two oxygen atoms make equal but opposite contributions to the X-ray emission, as Kanai *et al.* explain. The suppression of the harmonic emission occurred when the two oxygen atoms in the molecule are separated by exactly one complete wavelength of the electron wave. The soft X-ray emission amplitude from each of the two oxygen atoms is then exactly out of phase, leading to destructive interference in the total emission.

In principle, laser-driven electron-wave interference provides a powerful tool for measuring the distances between atoms in molecules with a resolution of a fraction of a nanometre. The technique directly pinpoints the position of the atomic nuclei and complements tomographic reconstruction⁴, which can retrieve the full electronic wavefunction. The information retrieved through tomographic imaging is more complete. But Kanai and co-workers' technique is much easier to implement as it requires only the measurement of the angular dependence of ion and harmonic signals, and requires no sophisticated processing of the data. By observing the emitted soft X-ray spectrum over a range of photon energies, it is possible to see the interference arising from different bond lengths.

This proof-of-principle experiment has only looked at a very simple, linear triatomic molecule, and it will have to be tested with complex molecules. Nevertheless, it may be possible to use this method to carry out pump-probe studies that follow the internuclear separation as it changes on a timescale of a few femtoseconds.

Jonathan P. Marangos is in the Blackett Laboratory, Department of Physics, Imperial College London, London SW7 2BW, UK.
e-mail: j.marangos@imperial.ac.uk

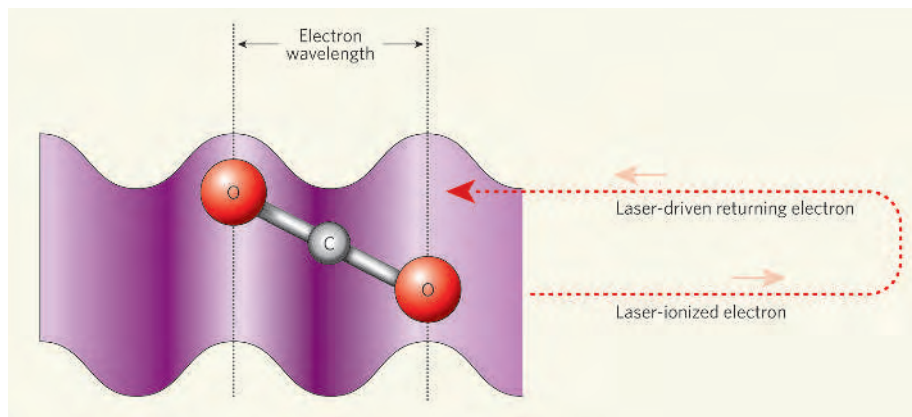


Figure 1 | The electron-wave interference technique. An electron escapes from a CO_2 molecule upon ionization by an intense near-infrared laser field. The laser field then accelerates and returns the emitted electron, producing soft X-ray radiation (not shown) when the electron recombines with the molecule — all within the same optical cycle. The returning electron has a wavelength given by $\lambda = h/p$ (where h is Planck's constant and p is the momentum of the electron). A strong interference modulation of the soft X-ray intensity is observed if λ matches the spacing between the two oxygen atoms (red) in the molecule.

1. Kanai, T., Minemoto, S. & Sakai, H. *Nature* **435**, 470–474 (2005).
2. Lein, M., Hay, N., Velotta, R., Marangos, J. P. & Knight, P. L. *Phys. Rev. A* **66**, 023805 (2002).
3. Rosca-Pruna, F. & Vrakking, M. J. J. *Phys. Rev. Lett.* **87**, 153902 (2001).
4. Itatani, J. *et al.* *Nature* **432**, 867–871 (2004).



50 YEARS AGO

OBITUARY — Prof. Albert Einstein. My first contact with Einstein was in Vienna in September 1913... he lectured to the Physics Section on "Gravitation", and his lecture quite obviously impressed most of his hearers as the work of a master-mind. But it was clear in the discussion which followed that many German-speaking men of science were not yet converted to his ideas... Einstein remained smilingly unperturbed and said he was prepared to stand or fall by the results of an empirical examination of his predictions. He had not very long to wait, for when I translated his popular work on "Relativity" in 1920, I suggested to him that he might like to include an appendix on the experimental confirmation of the theory...

His world-wide and unsought fame undoubtedly reached its zenith with the confirmation of his predicted gravitational deflexion of light rays by Eddington and others in 1919... His first comments in Britain on the results of the solar eclipse expeditions were published at the request of *The Times*... Referring to this in a letter to me, he wrote: "It cannot do any harm, for, thank God, the solar eclipse and the theory of relativity have nothing in common with politics... I should like to utilize the favourable circumstances to contribute as much as possible towards the reconciliation of German and English colleagues."

Robert W. Lawson

From *Nature* 28 May 1955.

100 YEARS AGO

On Friday, May 12... Lord Avebury, on behalf of his fellow trustees, received from Mr. Andrew Carnegie the gift of the full-sized model of the skeleton of the gigantic American dinosaur known as *Diplodocus carnegii*, which has been mounted in the reptile gallery of the Natural History Branch of the British Museum... It is almost an appalling thought that the skeleton of a creature which lived at least several million years ago should have come down in such a marvellous preservation to our own day.

From *Nature* 25 May 1905.

PLANT BIOLOGY

Auxin action

Judy Callis

Farmers and gardeners have long taken advantage of the growth-altering properties of the plant hormone auxin. The discovery of the elusive auxin receptor hints at how plant cells 'sense' and respond to this protein.

Even after 125 years of research in plant biology, we cannot answer the puzzle posed by this jaunty children's nursery rhyme:

*Oats and beans and barley grow,
Oats and beans and barley grow,
Can you, or I, or anyone know,
How oats and beans and barley grow?*

One reason is that plant growth involves at least eight different classes of compounds to coordinate growth, development and responses to the environment. One class, the auxins — from the Greek *auxein*, meaning to grow — stands out. Auxins are believed to be essential for plant life because, to date, no plant unable to synthesize auxin has been found¹. These proteins regulate many developmental programmes in the plant through their effects on cell growth, cell division and cell specialization. Yet how target cells 'sense' the presence of auxins has been unclear because, despite

many years of effort, no direct receptor for auxin had been isolated. In this issue, Dhar-masiri *et al.*² and Kepinski and Leyser³ (pages 441 and 446) reveal a pathway for auxin perception that is mediated by a novel receptor.

The enigma of how plants respond to auxin has held since 1880, when Charles Darwin reported a growth-stimulating substance that moved within plants⁴. In the 1930s, the substance was identified as indole-3-acetic acid (IAA)¹, which is now recognized as the most abundant of the auxins. But even without knowing how auxins work, people have long used them in agriculture as safe and effective agents for weed control, and in horticulture, for example to promote root development in cuttings (Fig. 1). In the case of weed control, the adage 'too much of a good thing' applies — too much auxin and plants die.

The identification of a plant auxin-binding protein (ABP1) 20 years ago marked a major

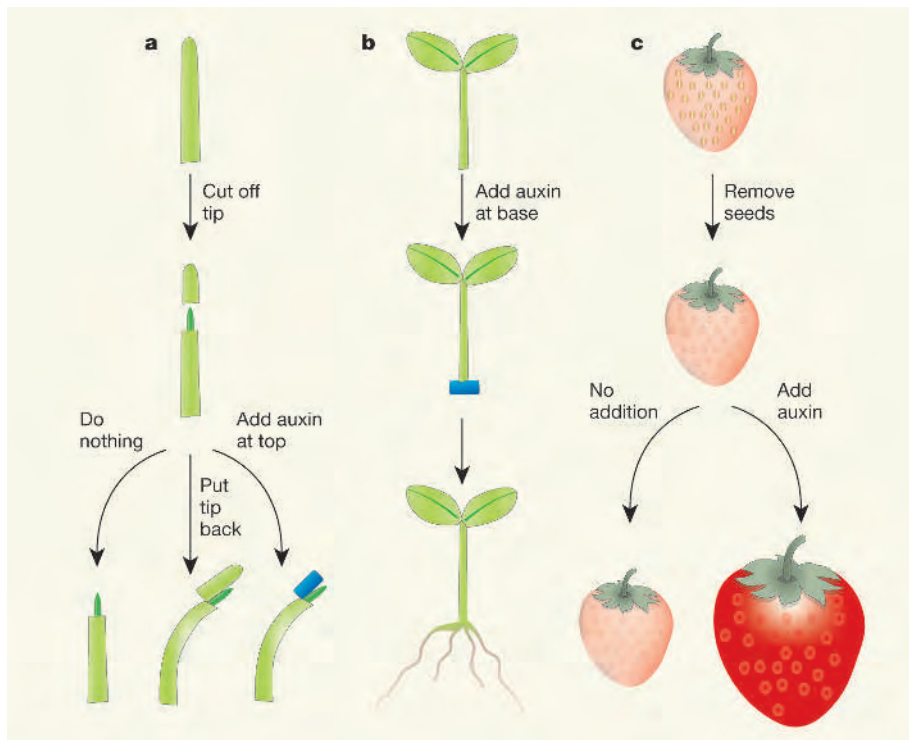


Figure 1 | Shoots, roots and fruits — the effects of auxin on plant growth. **a**, Grass seedlings have a sheath called the coleoptile that surrounds the first set of leaves. Growth of the coleoptile depends on the tip, and removal of the tip stops growth. Adding the tip back asymmetrically illustrates that the growth-promoting effect travels downward and not laterally, causing the seedling to bend because one side is growing faster than the other. Auxin can replace the tip for this effect. **b**, Stem cuttings can be induced by auxin to produce roots. **c**, Strawberries depend on auxin produced by their developing seeds for expansion and maturation. If the seeds are removed, little growth occurs. Normal growth can be restored with auxin.

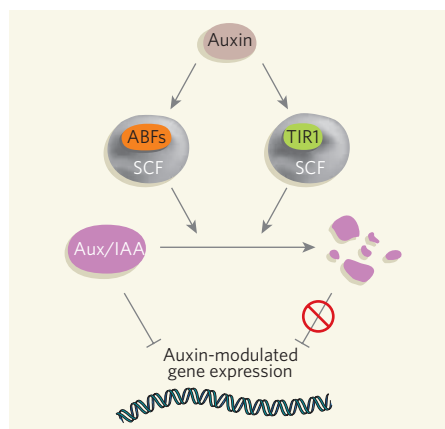


Figure 2 | Model of auxin action. Auxins act directly with SCF complexes containing either transport inhibitor response protein 1 (TIR1) or the related auxin-binding factors (ABFs). This catalyses the destruction of Aux/IAA proteins, which directly inhibit the genes that carry out the auxin response. The inhibitory effect of Aux/IAA is thus relieved, allowing auxin responses to occur.

advance in understanding auxin perception in plants⁵. Developing plants that lack ABP1 show defective cell elongation, fail to organize the basic plant body plan, and subsequently degenerate⁵. However, cell division still occurs in these plants, indicating that an auxin pathway to regulate cell division is still working. Finally, after ardent searching, comes the discovery of a surprising mode of auxin perception^{2,3}. The key proteins involved have been studied for several years, but they now show a direct and unexpected ability to 'sense' auxin.

Auxins cause rapid changes in gene expression, and two families of proteins have been identified in this response: auxin response factors (ARFs) and Aux/IAA proteins⁶. The genes encoding Aux/IAA proteins initially seemed to fit the mould of 'early response genes', because their expression increased rapidly following exposure to auxin. According to this model, Aux/IAA proteins would modulate the expression of 'late response genes' that encode factors directly involved in cell division and growth. But it turned out that these proteins actually suppress auxin-induced gene expression⁷, and that high auxin accelerates their destruction^{8,9}. So, rather than functioning solely as positive downstream messengers, Aux/IAA proteins also act as negative regulators, and their abundance must decline, at least initially, if auxin is to be sensed. How, then, does auxin mediate changes in Aux/IAA protein destruction?

The 'transport inhibitor response 1' protein (TIR1), previously shown to be a part of this response pathway, is now confirmed as the vital link^{2,3}. TIR1 was previously isolated in a genetic screen looking for mutant plants that show an altered response to auxin. The function of TIR1 was suggested by the presence of an 'F-box' motif in its protein sequence. This short sequence of amino acids is found in proteins that form part of a protein complex called the SCF, named after the first three of its

subunits to be identified: SKP1, cullin and F-box protein. This complex catalyses the covalent addition of a ubiquitin molecule to proteins, targeting them for destruction. The ubiquitin pathway is highly conserved among species, being found in all animals, fungi and plants — only bacteria lack it. Thorough biochemical work established that TIR1 was indeed part of a plant SCF complex¹⁰ that interacts with Aux/IAA proteins⁸. But how does auxin communicate with the SCF complex to alter Aux/IAA destruction? None of the precedents established for the regulation of SCF activity in animal or fungal systems seemed to apply.

Dharmasiri *et al.*² and Kepinski and Leyser³ show that auxin simply binds to TIR1 in the presence of an Aux/IAA protein, and somehow changes TIR1 activity. When plant TIR1 is made in animal cells, such as insect² or frog (*Xenopus*)³, auxin still binds to TIR1 in the presence of Aux/IAA, implicating TIR1, or TIR1 together with the Aux/IAA protein, as the sole components required to sense auxin by this pathway. Auxin perception is thus unique among SCF-mediated pathways. In the canonical SCF pathway, covalent modification of the target protein promotes its interaction with the SCF complex. In the auxin response, no target modification has been observed. Instead, it is the alteration of

the TIR1-containing SCF complex by non-covalent auxin binding that is responsible for the increased destruction of Aux/IAA proteins (Fig. 2).

Much remains to be done. The biochemical nature of the SCF–auxin–Aux/IAA interaction and the relationship between auxin perception by SCF and by ABP1 clearly merit further examination. Can the interaction of auxins with SCF account for the myriad auxin responses in growth and development? Let us hope that we will soon be able to answer the rhyme: yes, you and I can know how oats and beans and barley grow!

Judy Callis is in the Section of Molecular and Cellular Biology, University of California, 1 Shields Avenue, Davis, California 95616, USA. e-mail: jcallis@ucdavis.edu

1. Woodward, A. W. & Bartel, B. *Ann. Bot.* **95**, 707–735 (2005).
2. Dharmasiri, N., Dharmasiri, S. & Estelle, M. *Nature* **435**, 441–445 (2005).
3. Kepinski, S. & Leyser, O. *Nature* **435**, 446–451 (2005).
4. Darwin, C. *The Power of Movement in Plants* (Murray, London, 1880).
5. Napier, R. M., Davis, K. M. & Perrot-Rechenmann, C. *Plant Mol. Biol.* **49**, 339–348 (2002).
6. Hagen, G. & Guilfoyle, T. *Plant Mol. Biol.* **49**, 373–385 (2002).
7. Tiwari, S. B., Wang, X. J., Hagen, G. & Guilfoyle, T. J. *Plant Cell* **13**, 2809–2822 (2001).
8. Gray, W. M., Kepinski, S., Rouse, D., Leyser, O. & Estelle, M. *Nature* **414**, 271–276 (2001).
9. Zenser, N., Ellsmore, A. & Callis, J. *Proc. Natl Acad. Sci. USA* **98**, 11795–11800 (2001).
10. Gray, W. M. *et al. Genes Dev.* **13**, 1678–1691 (1999).

PARTICLE PHYSICS

Electrons are not ambidextrous

Andrzej Czarnecki and William J. Marciano

The best low-energy measurement yet obtained of the electroweak mixing angle — a central parameter of the standard model of particle physics — is the last hurrah for Stanford's powerful two-mile linear accelerator.

Like speeding rifle-bullets, high-energy electrons can spin about the direction of their motion. Electrons spinning clockwise are said by convention to be left-handed; those spinning anticlockwise are right-handed. Researchers on experiment E158 at the Stanford Linear Accelerator Center (SLAC) in California have announced¹ that, in contrast to bullets, the probability of an electron hitting a target depends slightly on whether its spin is right- or left-handed. This is a manifestation of a phenomenon known as parity violation — the difference between a fundamental interaction and its mirror image.

Most physical theories — from Newton's mechanics and Maxwell's electrodynamics, to relativity and quantum mechanics — are perfectly symmetric with respect to left–right interchange, or parity. But in 1956, Lee and Yang² suggested that the so-called weak interaction, responsible for the β decay of unstable nuclei and the feeble force of the accompanying

neutrinos, might exhibit a left–right preference — a feature soon established experimentally. In fact, the weak interaction was found to be maximally parity violating: only left-handed electrons participate in such reactions.

Through the work of Sheldon Glashow, Abdus Salam and Steven Weinberg³, the weak interaction was later unified with electromagnetism in what is now called the standard model of 'electroweak' phenomena. In this theory, electric and magnetic forces are mediated by massless photons, and parity is conserved: photons do not distinguish the handedness of electrons or any other elementary particles. Because electromagnetism dominates almost all physical, chemical and biological phenomena, the parity violation of weak interactions is not evident in the world around us. The very short-range weak force, on the other hand, is mediated by heavy, electrically charged analogues of the photon — W bosons — that interact with the left-handed

components of particles, but not with right-handed components. Thus, the maximal parity violation of weak interactions is easily accommodated. But why nature chose to violate parity remains a mystery, as do possible connections with the evolution of life (Box 1).

The unified electroweak theory also predicted the existence of another heavy, but neutral particle called the Z boson. Like the W, the Z also mediates weak interactions and violates parity. But, unlike that in W-mediated interactions, the degree of parity violation in Z-mediated interactions is not maximal, because of mixing effects that can be expressed in terms of the 'weak mixing angle', θ_W . And because the Z boson, like the photon, is electrically neutral, it also interferes with electromagnetic reactions, resulting in a small — usually unobservable — degree of parity violation in these processes.

The first definitive observation of a parity-violating effect caused by the Z boson — measured as $\sin^2\theta_W$, at an uncertainty of around 10% — was made some 30 years ago by the E122 experiment, a forerunner of E158 at SLAC⁴. E122 measured the difference between the scattering of left- and right-handed polarized electrons on a deuterium target; the observed asymmetry was consistent with theoretical expectations and was a historic confirmation of the standard model. Nowadays, W and Z bosons are routinely created at high-energy accelerators, and their properties have been thoroughly scrutinized — experiments at SLAC and at CERN in Geneva have used colliding electron and positron beams tuned to the Z boson mass at around 91 GeV

(9.1×10^{10} electronvolts) to measure the weak mixing angle, appropriate for high-energy studies, to an accuracy of less than 0.1% at high energies (Fig. 1).

Like its E122 predecessor, the E158 experiment¹ used the two-mile-long linear accelerator at SLAC — currently the world's highest-energy electron accelerator — to accelerate left- and right-handed electrons separately to about 50 GeV and scatter them off other electrons in a liquid-hydrogen target. This allowed the number of interactions between left-handed accelerated electrons and left-handed target electrons to be compared with the analogous number of right–right scattering events — no mean feat, as the average deflection of an electron after collision was less than 0.3°.

The observation of an asymmetry in event numbers at E158 — at an expected level of about 1.5 events in every 10 million scattered electrons — would mark the first discovery of parity violation in a simple electron–electron system and provide a novel determination of θ_W . The E158 researchers, led by spokesmen Emlyn Hughes (Caltech), Krishna Kumar (University of Massachusetts), Paul Souder (Syracuse University), and the analysis leader Yuri Kolomensky (University of California, Berkeley), having recorded a staggering 10^{16} electron–electron collisions, did indeed observe an asymmetry of the anticipated magnitude, a technical tour de force. The resulting experimental constraint, effective at relatively low energies (eff), of $\sin^2\theta_W^{\text{eff}} = 0.2397 \pm 0.0013$ (about 0.5% accuracy) provides the best existing determination of the weak mixing angle at low energy.

The value obtained by E158 for the effective low-energy weak mixing angle is considerably larger than that obtained in experiments at higher energies, such as those that created real Z bosons at CERN and SLAC. But should this value always be the same? The answer is, 'not quite'. Quantum effects, particularly 'clouds' of quark–antiquark excitations surrounding the electrons at short distances, modify the mixing of the photon and Z boson. This causes the effective weak mixing angle to change slowly as a function of the energy scale probed, a phenomenon known as 'running' (Fig. 1). The value of that parameter is expected to decrease by about 3% as one goes from relatively low-energy phenomena to the 'Z pole' of the CERN and SLAC measurements, and then to start increasing again at higher energies owing to quantum excitations of the heavy, charged W bosons. (That change of direction is incidentally analogous to the source of asymptotic freedom in quantum chromodynamics, which won its discoverers the 2004 Nobel prize⁵.) Results from E158 (ref. 1) and an earlier experiment on caesium atoms⁶, when compared with the more precise higher-energy determination, nicely establish the predicted low-energy running.

The relatively good agreement between the E158 and caesium results on the one hand, and

Box 1 | Parity violation and life

Could fundamental parity violation have macroscopic consequences in our living world? Some organic molecules, such as enzymes and amino acids found in living organisms, are maximally optically active. When subjected to linearly polarized light, they will rotate the plane of polarization in only one direction; that is, they have a definite handedness. This is not fundamental parity violation, because molecules with the opposite handedness can be created artificially. But why don't they participate in life's processes? This preference of life-giving molecules for one type of handed structure is called homochirality. Although homochirality does not seem to have any direct connection to parity violation, some scientists⁸ have speculated that parity violation in the weak interactions may have played an early role in evolution through some as yet undetermined enhancement mechanism. Such ideas are provocative but controversial. Should large organic molecules be found in extraterrestrial samples, it would be interesting to check the handedness of their optical activity.

A.C. & W.J.M.

theory on the other, allows one to rule out or constrain 'new physics' appendages to the standard model. For example, some theories invoke additional 'Z' bosons that could also couple to electrons and provide an added source of parity violation. The lack of any apparent deviation from the expectations of the standard model implies that the Z', if it exists, must have a large mass that suppresses its effect — indeed, it would have to be at least ten times heavier than the known Z boson.

The glorious history of the SLAC two-mile accelerator, encompassing many great discoveries, such as the first conclusive evidence for quarks, is now drawing to a close. E158 is the last high-energy fixed-target experiment at SLAC: the accelerator is now being converted into the Linac Coherent Light Source, the world's highest-energy free-electron laser. This bright source of X-rays will, literally, shed new light on a variety of questions in condensed-matter physics and the life sciences. The results from E158 provide a precision test of the standard model and establish the predicted running of the weak mixing angle via quantum quark effects, a fitting end to a distinguished career. ■

Andrzej Czarnecki is in the Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1, Canada. William J. Marciano is at the Brookhaven National Laboratory, Upton, New York 11973, USA. He is currently on leave at the Enrico Fermi Institute, University of Chicago. e-mails: czar@phys.ualberta.ca; marciano@bnl.gov

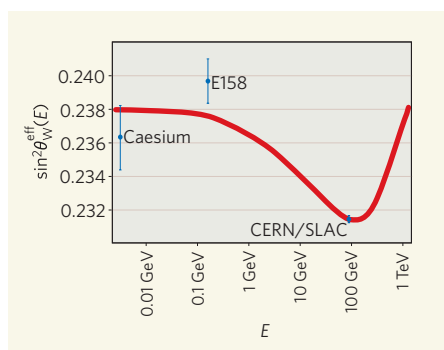


Figure 1 | Measurements of the weak mixing angle. The effective weak mixing angle θ_W^{eff} , the fundamental parameter of parity violation involving the Z boson, is expected to vary ('run') depending on the energy scale (E) probed by a given experiment. Using the very precise value obtained by high-energy studies at CERN and SLAC, the expected running at lower and higher energies is illustrated. Pictured at low energies is the new E158 result¹, as well as an earlier, somewhat less precise measurement of the weak mixing angle obtained from a study of parity-violating interference effects in atomic caesium⁶. Together, they nicely confirm the theoretically predicted running. (It should, however, be noted that data from a high-energy neutrino experiment at Fermilab⁷ suggest a significant, currently unexplained, deviation from the curve.)

1. Anthony, P. L. et al. (SLAC E158 collaboration) *Phys. Rev. Lett.* (submitted); preprint at <http://xxx.arxiv.org/abs/hep-ex/0504049> (2005).
2. Lee, T.-D. & Yang, C.-N. *Phys. Rev.* **104**, 254–258 (1956).
3. Glashow, S. L., Salam, A. & Weinberg, S. <http://nobelprize.org>.
4. Prescott, C. Y. et al. *Phys. Lett.* **B77**, 347–352 (1978).
5. Gross, D. J., Politzer, H. D. & Wilczek, F. <http://nobelprize.org>.
6. Wood, C. S. et al. *Science* **275**, 1759–1763 (1997).
7. Zeller, G. P. et al. *Phys. Rev. Lett.* **88**, 091802 (2002).
8. Cline, D. B. *Comments Nucl. Part. Phys.* **22**, 131–154 (1997).

BRIEF COMMUNICATIONS

Rare items often missed in visual searches

Errors in spotting key targets soar alarmingly if they appear only infrequently during screening.

Our society relies on accurate performance in visual screening tasks — for example, to detect knives in luggage or tumours in mammograms. These are visual searches for rare targets. We show here that target rarity leads to disturbingly inaccurate performance in target detection: if observers do not find what they are looking for fairly frequently, they often fail to notice it when it does appear.

Visual search is the subject of voluminous lab literature¹. Typically, observers perform several hundred searches and targets are presented on 50% of trials. But target prevalence in baggage checks or cancer screening is much lower (about 0.3% in routine mammography²).

We compared performance on high- and low-prevalence versions of an artificial baggage-screening task in which observers looked for 'tools' among objects drawn from other

categories. Semi-transparent objects were presented against noisy backgrounds and were sometimes overlapping; the number of objects in a display was 3, 6, 12 or 18; target prevalence was 1%, 10% or 50%. (For methods, see supplementary information.)

In the 1%-prevalence condition, 12 paid volunteer observers were each tested over 2,000 trials (broken into 250-trial blocks) that included only 20 target-present trials. Each observer was tested over 200 trials in the 10% and 50% conditions. Observers were given feedback on their performance, which included a points system designed to emphasize the importance of finding the target. Low-prevalence search has some similarity to vigilance tasks in which observers wait for fleeting signals^{3,4}; however, our search stimuli were continuously visible until observers chose to respond.

We measured error rates as a function of the number of objects (Fig. 1a). A prevalence of 50% produced 7% miss errors (failing to notice a target), which is typical for laboratory search tasks of this sort. However, errors increase dramatically and consistently as prevalence decreases: 10% prevalence produced 16% errors, and errors soared to 30% at 1% prevalence.

Errors were primarily 'misses'. False alarms (saying "yes" when targets are absent) were vanishingly rare (0.03%), despite incentives to produce the opposite response (see methods in supplementary information). Simply changing prevalence produced a fourfold increase in error rate. If similar effects occur in socially important searches, the implications are significant.

Why does this happen? The reaction-time data shown in Fig. 1b,c provide some clues. Observers require a threshold for quitting when no target has been found. This threshold is constantly adjusted: observers slow down after making mistakes and speed up after successes⁵. When targets are frequent, fast "no" responses will often lead to mistakes. As a result, "no" reaction times are slower than "yes" reaction times in a high-prevalence search (Fig. 1b). With infrequent targets, observers can successfully say "no" almost all of the time, driving down the quitting threshold. As seen in Fig. 1c, the result is a target-absent search that is too swift: observers abandon their search in less than the average time required to find a target.

The problem cannot be solved simply by adding pseudotargets to increase prevalence (for example, by asking baggage searchers to check for iPods as well as weapons). In a second experiment, we mixed common (44% prevalence), rare (10%), and very rare (1%) targets so that some target was present in 50% of trials. Here, observers missed just 11% of common targets, but 25% of rare targets and 52% of very rare targets (see supplementary information).

Is the prevalence effect just a by-product of a naive observer's unfamiliarity with the targets? In a separate investigation, we compared the miss-error rate for 4,000 trials at 1% prevalence (40 targets, 41% miss errors) with the miss-error rate for the first 100 trials at 34% prevalence (34 targets, 11% miss errors). It seems to be prevalence and not just the number of targets presented that is critical (see supplementary information).

Visual search is a ubiquitous human signal-detection task⁶. Heuristics that produce acceptable performance over a wide range of target prevalence may betray us at low prevalence. Because the experiments are burdensome, we do not have a clear idea whether these effects occur in the field^{7,8}. A scoring system in the laboratory cannot duplicate the motivation to find a gun or a tumour, nor the motivation to move the check-in line along. And the training of laboratory volunteers differs from that of professionals. Nevertheless, there are sufficient similarities between laboratory and field to indicate that we should find out whether the large increases in error demonstrated here also occur in socially important search tasks.

Jeremy M. Wolfe*†, Todd S. Horowitz*†, Naomi M. Kenner*

*Visual Attention Laboratory, Brigham and Women's Hospital, Boston, Massachusetts 02139, USA

†Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts 02114, USA
e-mail: wolfe@search.bwh.harvard.edu

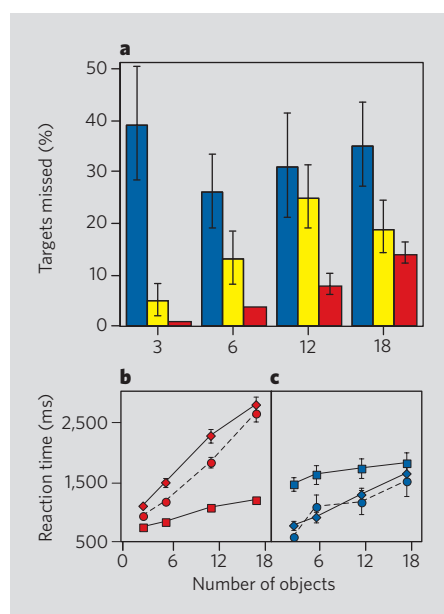


Figure 1 | The effects of target prevalence on search performance. **a**, Error rates for rare targets (blue bars, 1% prevalence), less rare targets (yellow bars, 10% prevalence) and common targets (red bars, 50% prevalence). Data from 12 observers are averaged. Error bars show $\pm$ s.e.m. for those 12 error rates. **b**, Reaction times for 50% prevalence. Typical reaction times are longer when the target is absent (circles) than when targets are present (squares). Diamonds show miss-error reaction times. **c**, Reaction times for 1% prevalence. At low prevalence, 'absent' responses are faster than 'present' responses. This leads to increased error rates. Error bars show $\pm$ s.e.m.

- Wolfe, J. M. in *Attention* (ed. Pashler, H.) 13–74 (Psychology Press, Sussex, UK, 1998).
- Gur, D. et al. *J. Natl Cancer Inst.* **96**, 185–190 (2004).
- Warm, J. S. in *Workload Transition: Implications for Individual and Team Performance* (eds Huey, B. M. & Wicken, C. D.) 139–170 (National Academy, Washington DC, 1993).
- Mackworth, J. *Vigilance and Attention* (Penguin, Harmondsworth, UK, 1970).
- Chun, M. M. & Wolfe, J. M. *Cogn. Psychol.* **30**, 39–78 (1996).
- Palmer, J., Verghese, P. & Pavel, M. *Vision Res.* **40**, 1227–1268 (2000).

7. Kundel, H. L. in *Medical Imaging 2000: Image Perception and Performance* (ed. Krupinski, E. A.) 135–144 (2000).
 8. Gur, D., Rockette, H. E., Warfel, T., Lacomis, J. M. & Fuhrman, C. R. *Acad. Radiol.* **10**, 1324–1326 (2003).

doi:435439a

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

BOSE-EINSTEIN CONDENSATES

Microscopic magnetic-field imaging

Today's magnetic-field sensors¹ are not capable of making measurements with both high spatial resolution and good field sensitivity. For example, magnetic force microscopy² allows the investigation of magnetic structures with a spatial resolution in the nanometre range, but with low sensitivity, whereas SQUIDS³ and atomic magnetometers⁴ enable extremely sensitive magnetic-field measurements to be made, but at low resolution. Here we use one-dimensional Bose–Einstein condensates in a microscopic field-imaging technique that combines high spatial resolution (within 3 micrometres) with high field sensitivity (300 picotesla).

Trapped cold atoms are ideal magnetic

sensors as they are very sensitive to changes in magnetic-field landscapes, even in the presence of large homogeneous offset fields. Density modulations in trapped thermal atomic clouds have already been used as a measure of magnetic field variation caused by irregular current flow in nearby conductors^{5–8}. We have produced a versatile, high-resolution sensor based on Bose–Einstein condensates (BECs). Its sensitivity is not limited by the temperature T of the cloud, but is rather determined by the chemical potential μ of the condensate, which can be orders of magnitude lower than $k_B T$ (where k_B is Boltzmann's constant).

The principles of the technique are shown in Fig. 1a. A BEC is trapped at the measurement site so that its density profile can be directly imaged. The spatially varying density is a measure of the potential energy and hence of the local magnetic-field variation. To probe spatial magnetic-field variations, we start by confining a one-dimensional BEC (in which μ is smaller than an energy quantum of transverse excitation)⁹ in an elongated magnetic micro-trap with strong transverse and weak longitudinal confinement, created by small conductors mounted on the surface of an atom chip¹⁰.

As a demonstration, we measured the magnetic-field variations above the 100- μm -wide current-carrying wire used to create the trap itself. Scanning the position of the BEC enabled us to reconstruct a full two-dimensional magnetic-field profile (Fig. 1b) near the wire with unprecedented accuracy (sensitivity of 4 nanotesla) at the measurement resolution (3 μm). From this map, we reconstructed the local current flow in the wire¹¹ and found extremely small angular deviations (2×10^{-4} root-mean-square radians) from a straight current path (for details, see supplementary information).

We also investigated an independent field landscape by placing a BEC close (5 μm) to a test wire structure. As long as this structure is grounded and carries no current, the atomic-density profile is homogeneous within the detection sensitivity. This corresponds to an upper bound in potential roughness of less than 10^{-14} eV, corresponding to a temperature of 200 picokelvin (field sensitivity of 300 picotesla). As soon as a small current (about 5 mA) is passed through the wire, a characteristic field profile is imaged.

The technique is applicable not only to magnetic fields, but can also be used to detect variations in electrostatic fields, as can be shown

by charging the probed structure electrically. This measurement is more sensitive than the one discussed above because the trapping parameters can be adjusted independently from the measured potential landscape by using a separate wire for holding the BEC.

The optimal potential sensitivity of a BEC used as a field sensor, $\Delta B = \gamma \Delta N / z_0^3$, is achieved if the trapping parameters are adjusted so that the cloud's transverse size matches the desired spatial resolution z_0 . Here ΔN is the minimal atom-number variation resolved by the imaging system, and γ contains all the atomic-physics parameters of the specific atom ($\gamma = 8.63 \times 10^{-29}$ tesla cubic metres for the ⁸⁷Rb used in our experiment). Currently available CCD (charge-coupled device) cameras allow atom-shot noise-limited detection with a ΔN value of better than 10 atoms per pixel in absorption imaging, so that a sensitivity, ΔB , of 1 nanotesla is possible even at a high spatial resolution of 1 μm (or 1 picotesla at 10 μm). By changing to a different atom with higher mass and/or by reducing the inter-atomic interaction, a significant increase in sensitivity can be achieved.

A comparison of different magnetic-field measurement techniques^{1–4,12} (see supplementary information) shows that BECs as magnetic sensors could reach unprecedented sensitivity over a large range of spatial resolution. The sample measurements we present here reach higher sensitivities than those obtained with established techniques operating at the same spatial resolution.

Stephan Wildermuth*, Sebastian Hofferberth*, Igor Lesanovsky*, Elmar Haller*, L. Mauritz Andersson*, Sönke Groth†, Israel Bar-Joseph†, Peter Krüger*, Jörg Schmiedmayer*

*Physikalisches Institut, Universität Heidelberg, Philosophenweg 12, 69120 Heidelberg, Germany e-mail: schmiedmayer@atomchip.org

†Department of Condensed Matter Physics, The Weizmann Institute of Science, Rehovot 76100, Israel

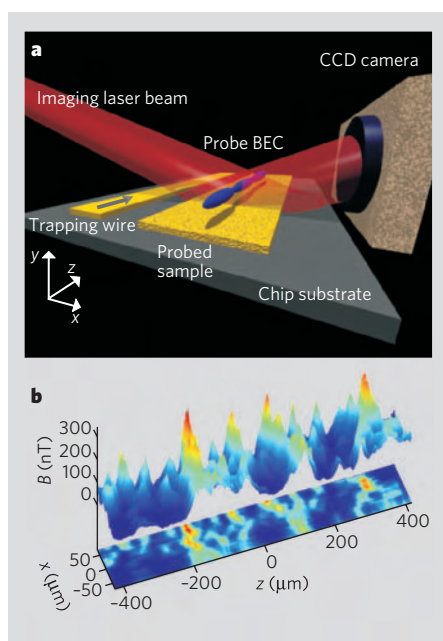


Figure 1 | One-dimensional Bose–Einstein condensate as a magnetic-field sensor.

a, Experimental set-up. A Bose–Einstein condensate is created and trapped by a current-carrying wire mounted on a silicon surface (atom chip) and is positioned above the sample to be probed. **b**, Two-dimensional scan of the magnetic landscape (field component along the wire direction) above a 100- μm -wide and 3.1- μm -tall gold wire. This profile has been reconstructed from 28 equally spaced one-dimensional atomic-density traces measured 10 μm above the current-carrying wire at a homogeneous offset field of 2 millitesla, so that relative field variations of only 4 p.p.m. stemming from slightly irregular current flow could be measured at a spatial resolution of 3 μm .

1. Bending, S. J. *Adv. Phys.* **48**, 499–535 (1999).
2. Freeman, M. R. & Choi, B. C. *Science* **294**, 1484–1488 (2001).
3. Faley, M. I. et al. *Supercond. Sci. Technol.* **17**, 324–327 (2004).
4. Kominis, I. K., Kornack, T. W., Allred, J. C. & Romalis, M. V. *Nature* **422**, 596–599 (2003).
5. Fortagh, J., Ott, H., Kraft, S., Günther, A. & Zimmermann, C. *Phys. Rev. A* **66**, 041604 (2002).
6. Esteve, J. et al. *Phys. Rev. A* **70**, 043629 (2004).
7. Leanhardt, A. E. et al. *Phys. Rev. Lett.* **89**, 040401 (2002).
8. Jones, M. P. A. et al. *Phys. Rev. Lett.* **91**, 080401 (2003).
9. Dunjko, V., Lorent, V. & Olshanii, M. *Phys. Rev. Lett.* **86**, 5413–5416 (2001).
10. Folman, R., Krüger, P., Schmiedmayer, J., Denschlag, J. & Henkel, C. *Adv. At. Mol. Opt. Phys.* **48**, 263–356 (2002).
11. Rous, P. J., Yongsunthorn, R., Stanishkevsky, A. & Williams, E. D. *J. Appl. Phys.* **95**, 2477–2486 (2004).
12. Oral, A. et al. *IEEE Trans. Magn.* **38**, 2438–2440 (2002).

doi:435440a

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

BRIEF COMMUNICATIONS ARISING online

♦ www.nature.com/bca see Contents pages.

WATER BEHAVIOUR

Glass transition in hyperquenched water?

Arising from: Y.-Z. Yue & C. A. Angell *Nature* **427**, 717–720 (2004)

It has been unclear whether amorphous glassy water heated to around 140–150 K remains glassy until it crystallizes or whether instead it turns into a supercooled and very viscous liquid. Yue and Angell¹ compare the behaviour of glassy water under these conditions to that of hyperquenched inorganic glasses, and claim that water stays glassy as it heats up to its crystallization point; they also find a ‘hidden’ glass-to-liquid transition at about 169 K. Here we use differential scanning calorimetry (DSC) heating to show that hyperquenched water deposited at 140 K behaves as an ultraviscous liquid, the limiting structure of which depends on the cooling rate — as predicted by theoretical analysis of the liquid-to-glass transition². Our findings are consistent with a glass-to-liquid transition-onset temperature (T_g) in the

region of 136 K (refs 3,4), and they indicate that measurements of the liquid’s properties may clarify the anomalous properties of supercooled water.

We hyperquenched micrometre-sized water droplets on a substrate held at 140 K (ref. 5) and immediately cooled it to 77 K at rates of 0.2, 2.0 and 5.0 K min^{−1}. DSC scans recorded subsequently (Fig. 1) show that the height of the endothermic peak (ΔC_p) decreases with increasing cooling rate. This effect disappears in DSC scans of samples that were prepared and cooled in the same manner but which were also annealed at 130 K (Fig. 1).

The mean ΔC_p (T_g) values of unannealed samples after cooling are: at 0.2 K min^{−1}, 1.7 ± 0.3 J K^{−1} mol^{−1} (135 ± 1 K) (from 18 samples); at 2.0 K min^{−1}, 1.1 ± 0.2 J K^{−1} mol^{−1} (136 ± 2 K) (from 9 samples); and at 5.0 K min^{−1}, 0.7 ± 0.1 J K^{−1} mol^{−1} (135 ± 1 K) (from 9 samples). For $T_g \approx 136$ K, water relaxes during deposition at 140 K for 16 min, moving towards metastable equilibrium.

The limiting structure obtained on subsequent cooling may be characterized in terms of a limiting ‘fictive’ temperature (T_f), which decreases with decreasing cooling rate². Decreasing T_f is experimentally observable by DSC on subsequent reheating, and is evident mainly as an increasingly pronounced overshoot² (Fig. 1); an overshoot can also develop upon annealing below T_g (ref. 6).

The ΔC_p increase of annealed samples⁴ (Fig. 1) contains an overshoot contribution, and water’s ‘true’ ΔC_p increase at T_g must be lower, approaching the value of about 0.7 J K^{−1}

mol^{−1} obtained on cooling at 5.0 K min^{−1}. A lower ΔC_p value seems consistent with increasingly ‘strong’ behaviour of supercooled water^{7–10}. Our findings therefore support the postulated fragile-to-strong transition of liquid water on cooling from ambient temperature into the supercooled and glassy state^{8,10}.

Our results are not consistent with the sub- T_g or “shadow” peak postulated by Yue and Angell, because their criterion is that the onset temperature of the peak is the same as the annealing temperature (see Fig. 3b in ref. 1). This is not observed here because T_g does not vary with annealing temperature (Fig. 1).

Ingrid Kohl*, Luis Bachmann†, Erwin Mayer*, Andreas Hallbrucker*, Thomas Loerting*†

*Institute of General, Inorganic and Theoretical Chemistry, and †Institute of Physical Chemistry, University of Innsbruck, 6020 Innsbruck, Austria
e-mail: ingrid.kohl@uibk.ac.at

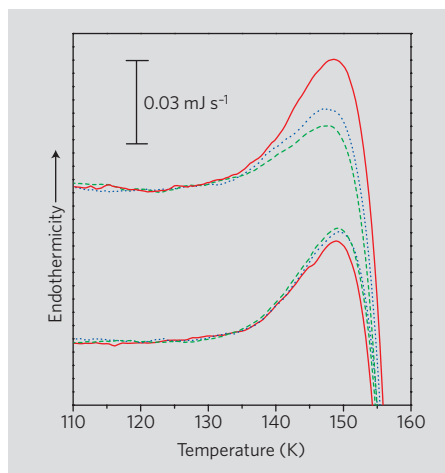


Figure 1 | Liquid-like relaxation in hyperquenched water at or below 140 K. Top curves: effect of cooling rate of unannealed hyperquenched water samples, after deposition at 140 K for 16 min, on differential-scanning calorimetry (DSC) measurements recorded on subsequent heating at 30 K min^{−1}; the cooling rate was increased from 0.2 K min^{−1} to 2.0 K min^{−1} and to 5.0 K min^{−1}, and the corresponding scans are indicated by solid, dotted and dashed lines, respectively. Note the decrease in the height of the endothermic step with increasing cooling rate. Bottom curves: effect of annealing at 130 K for 90 min on hyperquenched water samples, after deposition at 140 K for 16 min and cooling at 0.2, 2.0 and 5.0 K min^{−1}, on DSC scans recorded on subsequent heating at 30 K min^{−1} (line designations as in top curves). Note the disappearance of the effect of cooling rate for unannealed samples. Scans are normalized with respect to the weights and ice impurity of the samples and are drawn on the same scale. The ΔC_p values are corrected for 22% ice impurity⁵. The ordinate scale is for 1 mg sample weight. The scans are superimposed at low temperatures.

WATER BEHAVIOUR

Yue & Angell reply

Reply to: Kohl, I. *et al. Nature* doi:10.1038/nature03707 (2005)

The alternating support for and denial of a glass transition for amorphous water at 136 K has resumed after a hiatus of 20 years, during which it seemed secure. We revived the alternative interpretation¹ by looking again at the calorimetric signal that previously provided the most direct evidence for the glass transition^{2,3} — and now Kohl *et al.*⁴ present new data to support the original interpretation. We show here that their results are also consistent with our conclusions.

The new data of Kohl *et al.* show that if the exceedingly weak endothermic step (or peak),

originally reported as being the glass-to-liquid transition (T_g) of water², is a primary (glass-like) relaxation, then it is even weaker than previously supposed³ — only 3% above vibrational background and just a quarter the strength of the phenomenon that occurs in silica (SiO₂; the ‘strongest’ liquid known⁵). If it is a glass transition, then it is the broadest on record for a single-component system, with $\Delta T_g/T_g = 0.11$ (refs 3,5). As Kohl *et al.* observe, these characteristics would support the idea that water has undergone, during the hyperquench, a fragile-to-strong transition⁵, for

1. Yue, Y.-Z. & Angell, C. A. *Nature* **427**, 717–720 (2004).
2. Moynihan, C. T., Easteal, A. J., DeBolt, M. A. & Tucker, J. *J. Am. Ceram. Soc.* **59**, 12–16 (1976).
3. Johari, G. P., Hallbrucker, A. & Mayer, E. *Nature* **330**, 552–553 (1987).
4. Hallbrucker, A., Mayer, E. & Johari, G. P. *Phil. Mag. B* **60**, 179–187 (1989).
5. Kohl, I., Mayer, E. & Hallbrucker, A. *Phys. Chem. Chem. Phys.* **2**, 1579–1586 (2000).
6. Hodge, I. M. & Berens, A. R. *Macromolecules* **15**, 762–770 (1982).
7. Angell, C. A. *J. Non-Cryst. Solids* **102**, 205–221 (1988).
8. Ito, K., Moynihan, C. T. & Angell, C. A. *Nature* **398**, 492–495 (1999).
9. Debenedetti, P. G. *Metastable Liquids* (Princeton Univ., New Jersey, 1996).
10. Starr, F. W., Angell, C. A. & Stanley, H. E. *Physica A* **323**, 51–66 (2003).

doi:10.1038/nature03707

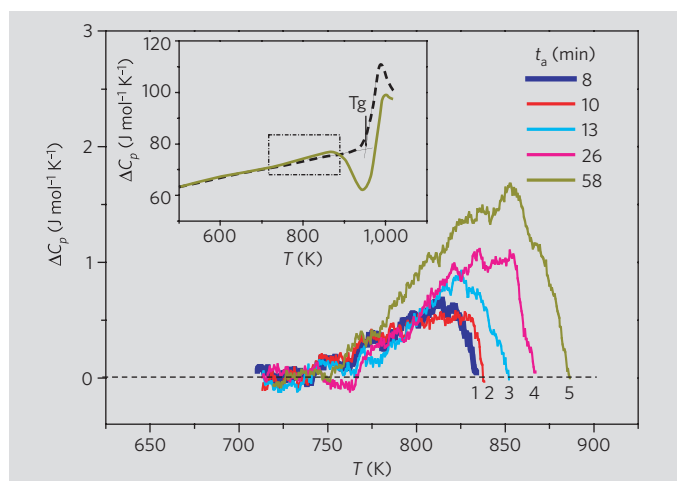


Figure 1 | Results analogous to those in Fig. 1 (top curves) of Kohl *et al.*⁴, but for non-crystallizing basaltic hyperquenched glass. Excess heat capacities over those of 'standard glass' (inset: standard glass, dashed line; annealed hyperquenched glass, solid line; boxed area, pre-peak) are shown. The temperature (T_g) of the glass-to-liquid transition for standard glass is 941 K. To demonstrate the analogy, we chose 823 K as the equivalent of the 140-K quench temperature of Kohl *et al.* for water ($823/941 = 140/160$). A sample was annealed at temperature T_a of 823 K for time t_a of 8 min to simulate the integral annealing of their sample over the 16-min deposition time, and then cooled to low temperature at 5 K min^{-1} , as in their fastest cool. Curves from the differential scanning calorimetry upscans: 1, curve from simulation described above; 2, curve obtained after 2 min of further annealing at 823 K, then fast cooling before the 30 K min^{-1} upscan, which yields the same upscan as does the protocol of Kohl *et al.*; 3–5, curves generated by simulating the slower cool-downs of Kohl *et al.* from 140 K and increasing anneal times at 823 K. If our measurements were curtailed by crystallization (at 850 K here), as for water, the diagram would look the same as that of Kohl *et al.*⁴.

which there is striking evidence^{6–8}.

However, we have to point out that the enhancement of the endothermic peak on slow cooling from 140 K, used by Kohl *et al.* to support the primary-relaxation scenario, is also the behaviour expected for an annealing pre-peak (our Fig. 1, inset), which was how we interpreted the weak endotherm¹. After all, the simplest way of understanding the existence of annealing pre-peaks (or "shadow" glass transitions) is to regard them as the annealing-enhanced glass transitions of the short-relaxation-time components ("microglasses"⁹) of the non-exponentially relaxing macroglass.

Our analogous scans are shown in our Fig. 1. If these were cut off by crystallization at a temperature of 830 K, in the same way as those for hyperquenched glassy water are cut off by crystallization at 155 K, then our Fig. 1 would have the appearance of Fig. 1 of ref. 4. The maximum ΔC_p ($0.6 \text{ J K}^{-1} \text{ mol}^{-1}$) for curve 1 is close to that ($0.7 \text{ J K}^{-1} \text{ mol}^{-1}$) for the 5-K min^{-1} scanned hyperquenched water of Kohl *et al.*⁴. The inset to Fig. 1 shows how weak these pre-peaks are relative to the real glass transition.

The scans in Fig. 1 suggest that the ' T_g ' for our system is 760 K, which is 20% below the real T_g of 941 K (see our Fig. 1, inset) and well below the annealing temperature of 823 K. We do not regard the identity of T_g with the annealing temperature as a central criterion, as Kohl *et al.* assert: neither our Fig. 1 here, nor Figs 2 and 3 in ref. 1, show this behaviour

(although others¹⁰ use this criterion). Furthermore, if we anneal for 55 days (rather than for minutes) at the lower temperature of 773 K before scanning, the pre-peak onset moves up to 920 K (Y.-Z. Y. *et al.*, unpublished results) and approaches the strength of the standard glass transition. The pre-peak onset temperature can evidently occur anywhere, depending on the fraction of the quenched-in energy that is left unrelaxed by the anneal.

In an earlier study¹¹, it was the magnitude of the unrelaxed enthalpy remaining when crystallization occurred that showed that T_g for water had been wrongly assigned. When there is no unrelaxed enthalpy in a substance undergoing vitrification, then annealing, or slow cooling, should result in T_g being raised¹² (an extreme being the case of silver-salt glasses after 5 years' annealing¹³). In Fig. 1 of Kohl *et al.*, we see no increase in T_g resulting from the slower cooling of the hyperquenched glassy water collected at 140 K.

We conclude that the new measurements of Kohl *et al.* leave unresolved the problem of water's post-annealing endothermic step (peak) at 136 K. It is possible that this conundrum could be solved by using their cool-and-anneal procedures^{3,4} on samples of water in nanoporous supports⁸ in which crystallization can be suppressed. We would expect the results to show that both of the above scenarios contribute to the effect.

An ironic twist to this problem is contained in findings¹⁴ that show that the T_g of the high-density polyamorph of water must be

135–140 K at ambient pressure, and that the transition (only partly seen) is strongly endothermic (large ΔC_p), as for our glass. On appropriate thermal treatment, this polyamorph should presumably show an interesting annealing pre-peak.

Yuanzheng Yue*, C. Austen Angell†

*Chemistry Section, Department of Life Sciences, Aalborg University, 9000 Aalborg, Denmark

†Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287, USA
e-mail: caa@asu.edu

1. Yue, Y.-Z. & Angell, C. A. *Nature* **427**, 717–720 (2004).
2. Johari, G. P., Hallbrucker, A. & Mayer, E. *Nature* **330**, 552–553 (1987).
3. Hallbrucker, A., Mayer, E. & Johari, G. P. *J. Phys. Chem.* **93**, 4986–4990 (1989).
4. Kohl, I., Bachmann, L., Mayer, E., Hallbrucker, A. & Loerting, T. *Nature* **435**, doi:10.1038/nature03707 (2005).
5. Ito, K., Moynihan, C. T. & Angell, C. A. *Nature* **398**, 492–495 (1999).
6. Starr, F. W., Angell, C. A. & Stanley, H. E. *Physica A* **323**, 51–66 (2003).
7. Faraone, A., Liu, L., Mou, C.-Y., Wen, C.-W. & Chen, S.-H. *J. Chem. Phys.* **121**, 10843–10846 (2004).
8. Maruyama, S., Wakabayashi, K. & Oguni, M. *Am. Inst. Phys. Conf. Proc.* **473**, 675–676 (2004).
9. Angell, C. A. *et al.* *J. Phys. Condens. Matter* **15**, S1051–S1068 (2003).
10. Johari, G. P. *J. Chem. Phys.* **119**, 2935–2937 (2003).
11. Velikov, V., Borick, S. & Angell, C. A. *Science* **294**, 2335–2338 (2001).
12. Moynihan, C. T., Easteal, A. J., DeBolt, M. A. & Tucker, J. *J. Am. Ceram. Soc.* **59**, 12–16 (1976).
13. Boehm, L., Ingram, M. D. & Angell, C. A. *J. Non-Cryst. Solids* **44**, 305–313 (1981).
14. Mishima, O. *J. Chem. Phys.* **121**, 3161–3164 (2004).

doi:10.1038/nature03708

The F-box protein TIR1 is an auxin receptor

Nihal Dharmasiri¹, Sunethra Dharmasiri¹ & Mark Estelle¹

The plant hormone auxin regulates diverse aspects of plant growth and development. Recent studies indicate that auxin acts by promoting the degradation of the Aux/IAA transcriptional repressors through the action of the ubiquitin protein ligase SCF^{TIR1}. The nature of the signalling cascade that leads to this effect is not known. However, recent studies indicate that the auxin receptor and other signalling components involved in this response are soluble factors. Using an *in vitro* pull-down assay, we demonstrate that the interaction between transport inhibitor response 1 (TIR1) and Aux/IAA proteins does not require stable modification of either protein. Instead auxin promotes the Aux/IAA–SCF^{TIR1} interaction by binding directly to SCF^{TIR1}. We further show that the loss of TIR1 and three related F-box proteins eliminates saturable auxin binding in plant extracts. Finally, TIR1 synthesized in insect cells binds Aux/IAA proteins in an auxin-dependent manner. Together, these results indicate that TIR1 is an auxin receptor that mediates Aux/IAA degradation and auxin-regulated transcription.

Since its discovery over 70 years ago, the plant hormone auxin or indole acetic acid (IAA) has been implicated in virtually every aspect of plant growth and development^{1,2}. In some tissues auxin regulates cell elongation, while in others the hormone promotes cell division. Recent studies also indicate that auxin acts as a morphogen during embryogenesis and in the root meristem^{3,4}. Despite the importance of auxin to the plant, many aspects of auxin signalling are poorly understood. In particular, the identity of the auxin receptor(s) is unknown. The best-characterized candidate receptor, the auxin binding protein 1 (ABP1), was identified by virtue of its auxin binding activity⁵. Although some characteristics of ABP1 are consistent with receptor function, the role of this protein in auxin signalling has not been determined.

In contrast, some aspects of auxin-regulated transcription are better understood. Two families of transcription factors, the auxin response factor (ARF) and Aux/IAA proteins, have been implicated in this process. The ARF proteins bind DNA directly and either activate or repress transcription depending on the ARF⁶. The Aux/IAA proteins exert their effects by binding to the ARF proteins through a conserved dimerization domain^{7,8}. At least in the case of the activating ARFs, the effect of Aux/IAA binding is to repress transcription.

Auxin regulates transcription by stimulating the degradation of the Aux/IAA proteins^{9–12}. Recent studies indicate that auxin acts by promoting an interaction between the Aux/IAA proteins and the ubiquitin protein ligase SCF^{TIR1} (ref. 10). In *Arabidopsis*, the Aux/IAA proteins are encoded by a family of genes comprised of 29 members⁸. Most of these proteins share four conserved regions designated domains I to IV. Domain II, including the conserved amino-acid residues GWPPV, has been implicated in the degradation of a luciferase reporter protein¹³ and interacts with SCF^{TIR1}, suggesting that domain II is the auxin degron^{14,15}. Mutations within domain II result in stabilization of the affected Aux/IAA protein and defects in auxin response^{10–12,16}.

How auxin promotes the interaction between Aux/IAA and SCF^{TIR1} is not known. Substrate recognition by many other cullin-based E3 ligases requires substrate modification, typically

phosphorylation, although proline hydroxylation and glycosylation have also been reported^{17–20}. In contrast, new evidence suggests that none of these mechanisms are likely to be involved in auxin-induced Aux/IAA degradation^{14,15}. Similarly, the recent suggestion that a parvulin-type prolyl isomerase may be involved in the interaction¹⁴ now appears unlikely with the discovery that the parvulin inhibitor juglone has a nonspecific effect on the Aux/IAA–SCF^{TIR1} interaction¹⁵. Instead, it has been suggested that substrate recognition requires an auxin-dependent modification of TIR1 or an associated protein, rather than the substrate¹⁵. To investigate this question, we adopted a biochemical approach based on the *in vitro* interaction assay we have previously described¹⁴. Here we report compelling evidence that auxin regulates degradation of the Aux/IAA proteins by binding directly to TIR1.

The TIR1–Aux/IAA interaction does not require a stable modification of either protein

We first asked whether the Aux/IAA–SCF^{TIR1} protein interaction is affected by temperature, reasoning that if stable modification of the TIR1 protein involves an enzyme, the reaction should be temperature-dependent. However, when pull-down assays were carried out at temperatures ranging from 4 to 25 °C, no differences in the recovery of TIR1–Myc were observed, suggesting that an enzyme-mediated modification is not involved (Fig. 1a). At 37 °C, the amount of TIR1–Myc in the pull-down decreased, possibly because the TIR1–Aux/IAA interaction is less stable at this temperature.

In a previous study, we showed that auxin promotes the interaction between TIR1 and the Aux/IAA protein within 5 min of addition to the pull-down reaction and that the response is saturated after 30 min (ref. 21). The results in Fig. 1b confirm that at both 0.5 and 50 µM 2,4-dichlorophenoxy acetic acid (2,4-D), the interaction is saturated after 20 min at 4 °C. To further explore the effects of temperature we performed pull-down experiments using 0.5 and 50 µM 2,4-D at 4 °C and 25 °C with an incubation time of 25 min. The results in Fig. 1c demonstrate that the kinetics of the response is similar at the two temperatures. Taken together, these results show

¹Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

that the effect of auxin is largely temperature-independent, arguing against the involvement of an enzyme-based modification of TIR1 or any other protein.

The auxin receptor co-purifies with TIR1

The pull-down assay that we use to examine the interaction between TIR1 and the Aux/IAA proteins involves extensive washes in buffer lacking auxin. Because our results suggest that a stable protein modification is not required for Aux/IAA binding, we wondered if the interaction might depend on the continuous presence of auxin. As shown in Fig. 1d, the addition of auxin into the washing buffer greatly enhanced the recovery of TIR1–Myc from the pull-down reaction, suggesting that auxin acts to stabilize the interaction between TIR1 and the Aux/IAA protein and that the hormone is required continuously for this effect.

Next we investigated whether partially purified TIR1–Myc is capable of interacting with IAA7 protein in an auxin-dependent manner. TIR1–Myc was immunoprecipitated from plant extracts in the presence or absence of 50 μ M 2,4-D using anti-Myc antibody linked to sepharose beads (Covance). After the beads were washed with several bed volumes of buffer, TIR1–Myc was eluted with 50 mM Tris-Cl (pH 7.2), 300 mM NaCl, 0.5% Tween and 10% dioxane. Eluted TIR1–Myc was then used in pull-down assays with GST–IAA7. The results in Fig. 2a show that immunoprecipitated TIR1–Myc interacts with GST–IAA7 in an auxin-dependent manner similar to TIR1–Myc in crude plant extracts. Auxin treatment of the plant extract before immunoprecipitation did not affect the subsequent interaction, providing additional support for the hypothesis that auxin does not stably modify TIR1 or a protein associated with

TIR1 (Fig. 2b). These results indicate that all the factors necessary for auxin-induced interaction, including the receptor, are present in the anti-Myc immunoprecipitate.

Because the Aux/IAA–TIR1 interaction depends on the continuous presence of auxin (Fig. 1d), it should be possible to purify TIR1–Myc by auxin-dependent binding to GST–IAA7. To investigate this possibility, pull-down assays were carried out with GST–IAA7 in the presence of auxin in both the pull-down buffer and the washing buffer. The TIR1–Myc protein was then eluted into washing buffer without auxin by vigorous agitation. Eluted TIR1–Myc was used in the pull-down reaction using GST–IAA7 in the presence or absence of auxin. TIR1–Myc recovered by this method was still responsive to auxin (Fig. 2c). Further, preincubation of GST–IAA7 with auxin either in the extraction buffer or in the plant extract did not affect the subsequent interaction in the pull-down assay (data not shown).

Auxin binds directly to SCF^{TIR1}

Because auxin does not appear to cause a stable modification of TIR1 or the Aux/IAA proteins, we hypothesized that auxin may regulate the interaction between Aux/IAA and SCF^{TIR1} by directly binding to the SCF^{TIR1} complex. To test this possibility, we carried out pull-down assays using GST–IAA7 and crude extracts from *GVG::TIR1-myc* seedlings in the presence of [³H]-IAA. After washing the GST–IAA7 beads extensively, radioactivity retained with the GST–IAA7 beads was measured by scintillation counting. The results in Fig. 3a show that GST–IAA7 retained [³H]-IAA, but GST–AXR2-1 did not. The AXR2-1 protein has an amino-acid substitution in domain II that prevents interaction with SCF^{TIR1} (ref. 10).

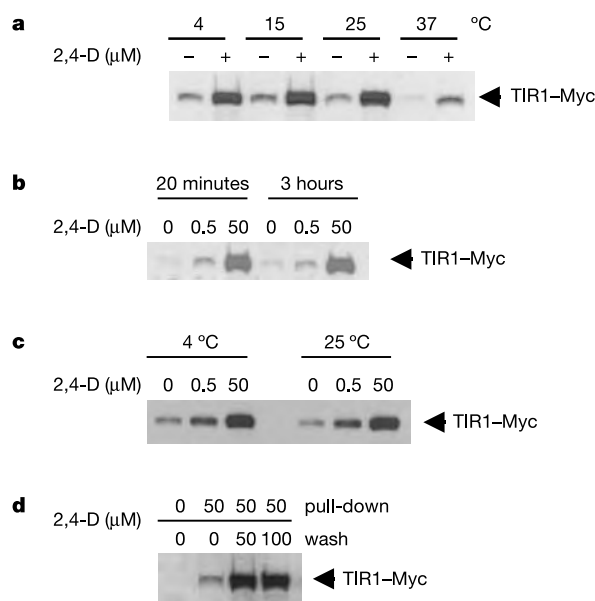


Figure 1 | Auxin-induced interaction between Aux/IAA and the SCF^{TIR1} is dependent on auxin concentration but not temperature.

Pull-down reactions were carried out using recombinant GST–IAA7 and crude plant extracts prepared from *tir1-1*, *GVG-TIR1-myc* seedlings. Protein bound to GST–IAA7 was washed, separated on SDS–PAGE and immunoblotted with anti-Myc antibody. **a**, Pull-down reactions were carried out at indicated temperatures for 20 min in the presence or absence of 50 μ M 2,4-D. **b**, Pull-down reactions were carried out at 4 °C for the indicated times with different 2,4-D concentrations. **c**, Pull-down reactions were incubated at the indicated temperatures and 2,4-D concentrations for 25 min. **d**, Pull-down reactions were incubated for 25 min at 4 °C and the glutathione beads were rinsed once with 1 ml of washing buffer without 2,4-D and then washed three times with washing buffer with or without 2,4-D.

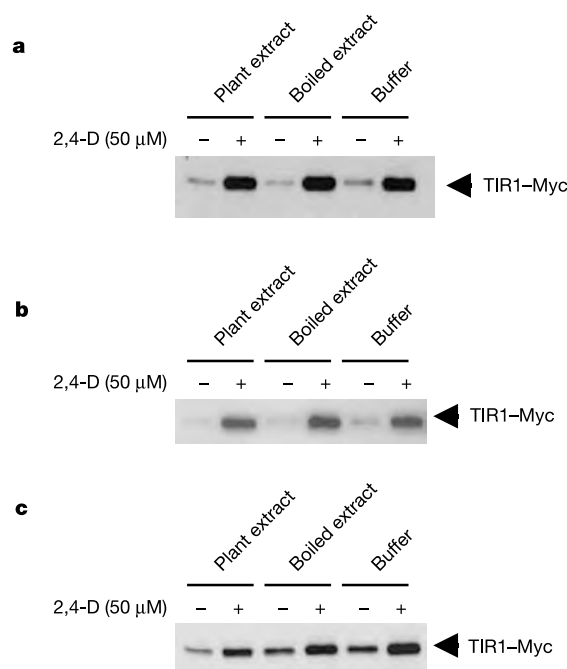


Figure 2 | Partially purified TIR1–Myc interacts with GST–IAA7 in an auxin-dependent manner.

a, TIR1–Myc was immunoprecipitated from crude plant extracts with anti-Myc antibody. Eluted protein was used in a pull-down assay as described in Methods. 100 μ g of wild-type plant extract (plant extract), boiled plant extract (boiled extract) or extraction buffer (buffer) was added to the pull-down reaction. **b**, Crude plant extract was incubated with 50 μ M 2,4-D for 1 h and then TIR1–Myc was immunoprecipitated with anti-Myc antibody before the pull-down assay as above. **c**, TIR1–Myc was first pulled-down with GST–IAA7, washed in the presence of 50 μ M 2,4-D and then eluted into washing buffer without 2,4-D. Eluted protein was used in the pull-down assay as described in **a**.

To explore further the biological relevance of this binding we performed competitive binding experiments with active auxins and related compounds. The active auxins IAA, 1-NAA and 2,4-D competed efficiently with [3 H]-IAA for binding while benzoic acid and tryptophan—related molecules with no auxin activity—did not compete effectively (Fig. 3b). These results suggest that the auxin receptor is localized to the SCF^{TIR1} complex. Scatchard analysis indicated that the apparent dissociation constant K_d of the receptor for IAA is 84 nM (data not shown). In addition, our results demonstrate that the natural auxin IAA has a higher affinity for the receptor than either 2,4-D or 1-NAA, consistent with our earlier results indicating that IAA promotes the interaction between Aux/IAA and TIR1 at nanomolar concentrations while 2,4-D and 1-NAA are active at micromolar levels¹⁴. The median inhibitory concentration IC_{50} for IAA, 1-NAA, and 2,4-D is 0.12, 1.3 and 1.4 μ M, respectively, in this assay. It should be noted however that the K_d and IC_{50} values were obtained in binding assays with crude plant extracts containing unknown levels of endogenous IAA. In addition, it is likely that a fraction of the exogenous IAA is metabolized during the course of the experiment. Although the IC_{50} values provide information on the relative affinity of each compound for the receptor, it will be necessary to determine definitive K_d and IC_{50} values using purified proteins.

Because auxin is recovered with GST-IAA7 in the presence of crude plant extract, we tested whether GST-IAA7 itself interacts with [3 H]-IAA in the extraction buffer. The results in Fig. 3c show that [3 H]-IAA is recovered with GST-IAA7, but the amount recovered is not affected by addition of excess cold IAA. A similar level of binding was observed with the AXR2-1 protein (data not shown). Thus IAA binding to GST-IAA7 is nonspecific, suggesting that the Aux/IAA proteins do not function as receptors.

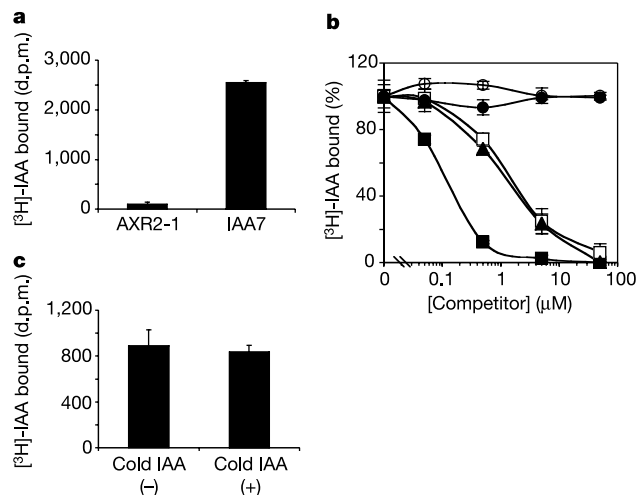


Figure 3 | [3 H] IAA interacts with the SCF^{TIR1} complex. **a**, Pull-down reactions were carried out with either GST-IAA7 or GST-AXR2-1 in the presence of 200 nM [3 H]-IAA. The retained fraction of [3 H]-IAA after washing the glutathione beads was measured by scintillation counting. Each point is the mean of three values $\pm$ standard deviation (s.d.). **b**, Competitive binding of [3 H]-IAA to SCF^{TIR1} in the presence of unlabelled IAA (black squares), 2,4-D (white squares) and 1-NAA (black triangles). The related compounds benzoic acid (white circles) and L-tryptophan (black circles) do not compete with [3 H]-IAA. Pull-down reactions were performed in the presence of 50 nM [3 H]-IAA and the indicated amount of competitor. Each assay was replicated three times and results were normalized relative to no competitor. Error bars represent s.d. **c**, GST-IAA7 was incubated in buffer containing [3 H]-IAA with or without cold IAA and recovered as described above. d.p.m., disintegrations per minute. Values are the mean of three experiments $\pm$ s.d.

TIR1 is an auxin receptor

SCF^{TIR1} contains Cullin homologue 1 (CUL1), arabidopsis skp1-like 1 (ASK1) and RING-box protein 1 (RBX1) in addition to TIR1 (ref. 19). Because these proteins are common to many different SCF complexes, it is likely that auxin interacts directly with either TIR1 or a protein tightly associated with TIR1. To distinguish between these two possibilities, we first synthesized TIR1 *in vitro* using a wheat germ extract (Promega). As shown in Fig. 4a, TIR1 synthesized in this way interacts with GST-IAA7 in an auxin-inducible manner. However, it is possible that endogenous proteins present in the wheat germ extract are facilitating this response. To eliminate this possibility we synthesized H6-TIR-Myc protein in insect cells. GST-IAA7 pull-downs were performed using extracts prepared from insect cells expressing H6-TIR-Myc. The results in Fig. 4b show that the TIR1-Myc in the extract interacts with GST-IAA7 in an auxin-dependent manner whereas GST-AXR2-1 does not. TIR1 and GST-IAA7 are the only plant-derived proteins in the pull-down assay, so this result implies that auxin binds directly to TIR1.

Because CUL1, ASK1, and RBX1 are highly conserved between plants and animals it is possible that TIR1 synthesized in wheat germ extract or insect cells associates with endogenous ASK1 (SKP1 in insects) and CUL1. To begin to investigate the importance of SCF assembly for TIR1 function, we synthesized TIR1 protein lacking 75 amino acids near the amino terminus including the F-box motif (Δ FB-TIR1). This protein did not respond to auxin (Fig. 4a) in a pull-down assay. There are a number of possible interpretations of this result. One is that the deleted sequences directly mediate IAA and/or Aux/IAA binding. Alternatively, assembly of TIR1 into an SCF, or at least interaction with ASK1 may be required for TIR1 function. Additional studies will be required to resolve this issue.

If TIR1 functions as an auxin receptor, the level of saturable IAA binding should be reduced in the *tir1* mutant. However, the amount of binding in *tir1* extracts is indistinguishable from the wild type (data not shown). One possible explanation for this result is that TIR1 is a member of a family of related receptors. The *TIR1* gene is in a small subfamily of seven related F-box protein genes²². Three of these genes, *At4g03190*, *At3g26810* and *At1g12820* have been named *AFB1*, *AFB2* and *AFB3* respectively, for AUXIN SIGNALING F-BOX protein. In a previous report *At4g03190* and *At3g26810* were designated *LRF1* and *LRF2* (ref. 23). An alignment of TIR1, the AFB proteins, and the related F-box protein COI1 is shown in Fig. 5a. COI1 is required for jasmonic acid response and is in the same subclade as TIR1 and the AFB proteins^{22,24}. To investigate the role of AFB1, AFB2 and AFB3 proteins in auxin binding, we performed GST-IAA7 pull-down experiments in the presence of [3 H]-IAA using quadruple mutant lines lacking TIR1, AFB1, AFB2 and AFB3. The results shown in Fig. 5b show that saturable [3 H]-IAA was

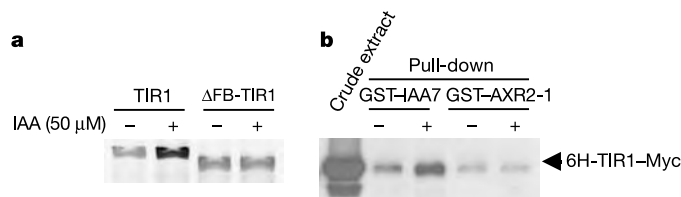


Figure 4 | TIR1 protein translated *in vitro* or expressed in insect cells interacts with GST-IAA7 in an auxin-dependent manner. **a**, TIR1 and Δ FB-TIR1 were translated *in vitro* in the presence of 35 S-methionine using the wheat germ system and directly used in pull-down assays in the presence or absence of 50 μ M IAA. **b**, 6H-TIR1-Myc was expressed in High Five insect cells as described in the Methods. Crude protein extracts were used in pull-down reactions with either GST-IAA7 or GST-AXR2-1 in the presence or absence of 50 μ M IAA.

undetectable in this line, indicating that auxin binding is dependent upon TIR1 and the AFB proteins.

Discussion

Our results provide compelling evidence that TIR1 is an auxin receptor that mediates rapid degradation of Aux/IAA proteins and consequent changes in expression of auxin-regulated genes. At this point it is not clear how auxin stimulates the interaction between SCF^{TIR1} and its substrates. It is possible that auxin binds to TIR1 and promotes a conformational change that favours Aux/IAA binding. Alternatively, auxin may bind cooperatively to both TIR1 and the Aux/IAA protein, thus stabilizing the SCF^{TIR1}-substrate complex. The site of auxin binding within TIR1 is also not known. The protein consists of the N-terminal F-box motif, a short spacer region of about 40 residues, 16 degenerate leucine-rich repeats (LRRs), and a C-terminal tail of approximately 70 residues (Fig. 5a). A comparison between TIR1, the three AFB proteins, and COI1 does not reveal any major TIR1/AFB specific domains that might be an auxin-binding pocket, although there are several short stretches of amino acids in the TIR1/AFB proteins that are not present in COI1. LRRs, which make up the bulk of TIR1, are typically thought to facilitate protein-protein interactions. However, brassinosteroids have recently been shown to bind to an LRR and adjacent sequences within the

brassinosteroid receptor BRI1²⁵. Thus it is possible that auxin binds one or more LRRs within TIR1. This and other possibilities will be assessed in future studies.

At this point it is not clear whether SCF^{TIR1} and related SCFs are the only targets of auxin action. In addition to changes in gene expression, auxin appears to regulate ion transport through the plasma membrane^{26–29}. One of the best-characterized auxin responses is a rapid increase in plasma membrane H⁺-ATPase activity, an effect that has been associated with cell elongation²⁹. Because this response is too rapid to be mediated by transcriptional changes, it probably does not directly require TIR1-dependent degradation of Aux/IAA proteins. It is possible that TIR1 promotes the degradation of other proteins that regulate ion transport. Alternatively, and perhaps more probably, activation of ATPases and other ion transporters may be controlled by a different auxin signalling pathway.

Members of the SCF family of E3 ligases are known to play important roles in many aspects of cellular regulation in eukaryotes^{17,19,30}. In addition, genomic analyses indicate that there are a large number of uncharacterized F-box proteins in both plant and metazoan genomes¹⁹. Among characterized SCFs, substrate recognition typically involves phosphorylation within the substrate degron¹⁹. Thus our discovery that the small ligand IAA regulates

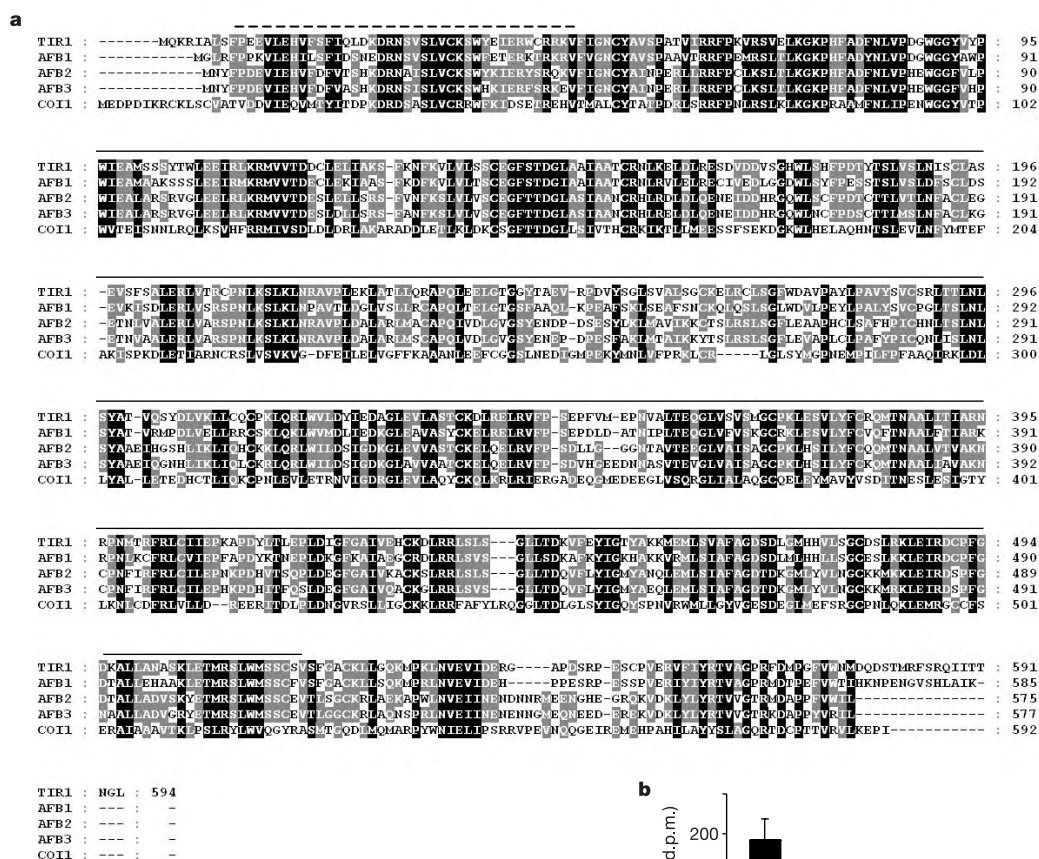


Figure 5 | A small family of F-box proteins contributes to auxin binding.
a, TIR1 was aligned with three closely related F-box proteins called AFB1, AFB2 and AFB3. The COI1 protein is the most closely related F-box protein that does not function in auxin response. The dashed line above the sequence indicates the position of the F-box motif, and the solid line

indicates the position of the LRRs. Residues shaded black are similar in all five proteins, while those shaded grey are similar in three of the proteins.
b, [³H]-IAA binding in the *Col* (wild type, WT) crude protein extracts and protein extracts from *tir1-1 afb1-1 afb2-1 afb3-1* (quadruple mutant, QD) seedlings. Values are the mean of three experiments $\pm$ s.d.

SCF^{TIR1}-substrate recognition represents a new paradigm for SCF regulation. Future studies will reveal exactly how IAA interacts with TIR1 and the Aux/IAA proteins and whether other plant and animal SCFs are also ligand regulated.

METHODS

Plant material. *Arabidopsis* (Col) *tir1*, *GVG::TIR1-myc* seedlings were grown under sterile conditions on vertically oriented ATS (*Arabidopsis thaliana* medium +1% sucrose) plates at 22 °C under constant light. Twelve-day-old seedlings were transferred into liquid ATS medium and TIR1-Myc expression was induced with 30 µM dexamethasone.

Pull-down assays. Total protein was extracted from seedlings in a buffer containing 50 mM Tris-Cl (pH 7.2), 100 mM NaCl, 10% glycerol, 1 mM PMSF, 10 µM MG132 and complete mini-protease inhibitors according to the manufacturer's instructions (Roche Diagnostics). Cell debris was removed by centrifugation at 10,000 g for 10 min. Total protein concentration was estimated by the Bradford assay (BioRad). Recombinant GST-IAA7 and GST-AXR2-1 were expressed in *Escherichia coli* and purified using glutathione beads according to standard protocols. GST-IAA7 or GST-AXR2-1 (3–4 µg) was incubated with 800 µg of total crude plant protein extract and incubated at 4 °C for 1 h unless otherwise specified. Glutathione beads were recovered by a brief centrifugation and washed three times with 1 ml of washing buffer (50 mM Tris-Cl (pH 7.2), 100 mM NaCl, 10% glycerol and 0.1% Tween 20. Where it is specified, 2,4-D was added to the washing buffer.

For *in vitro* translation, the *TIR1* coding sequence was cloned into *Xho1/Xba1* sites of the *pTNT* vector creating the plasmid *pNDS45*. To generate the Δ FB-TIR1 mutant, a *Kpn1* site was introduced to the *TIR1* complementary DNA using QuikChange site-directed mutagenesis kit (Stratagene) with primers 5'-AGC GAATGACCTTGTCTCGGTACCAGAAGAGGTACTAG-3' and 5'-CTCTAGTACCTCTCTGTTACCGACAAAGGCTATTCGCG-3'. The resulting plasmid, *pNDS46*, was digested with *Kpn1* to release a 228-base-pair (bp) *Kpn1* fragment that includes the F-box domain, but leaving the N-terminal 8-amino-acid residues intact. The rest of the plasmid was religated to create *pNDS47* encoding Δ FB-TIR1 lacking amino acids 9 to 84. The full-length and Δ FB-TIR1 proteins were expressed in the coupled wheat germ extract system (Promega) in the presence of ³⁵S-trans label (1,175 Ci mmol⁻¹; MP Biochemicals). For pull-down assays, 20 µl of translated product was incubated with 3–4 µg of GST-IAA7 in 200 µl extraction buffer for 5 h and unbound proteins were washed as described above. Pull-down mixtures were separated on SDS-polyacrylamide gel electrophoresis (PAGE) and the bound TIR1 was detected by using the Phosphorimager (Typhoon 9200, Amersham Biosciences).

Auxin binding assays. To determine auxin binding, pull-down assays were done as described above using plant extract containing TIR1-Myc, except that [³H]-IAA (specific activity 20 mCi mmol⁻¹) was added to the pull-down reaction. Each reaction had a final concentration of 50 nM [³H]-IAA unless otherwise specified. After washing three times in the presence of excess unlabelled IAA, glutathione beads were resuspended in 100 µl of water and mixed with scintillation fluid. The radioactivity of the bound [³H]-IAA was measured using a scintillation counter.

Expression of TIR1 in insect cells. The TIR-Myc coding sequence was amplified from *pGB2823* using the oligonucleotides 5'-CACCATGCAGAAGCGAATAGCCTTGTC-3' and 5'-AGCTTATCGATTTCGAACCCGGGGTAC-3' and first cloned into the *pENTR-D/TOPO* and then into the *pDEST10* using the Gateway system according to the manufacturer's instructions (Invitrogen). The resulting plasmid, *pDEST10-H6-TIR1-myc* was then transformed into *DH10Bac* competent cells. *E. coli* colonies with recombinant bacmid were identified according to the manufacturer's instruction (Invitrogen). Initial viral production and amplification were done using Sf9 cells. Bacmid DNA (1 µg) diluted in 200 µl of Sf900 II SFM medium (Invitrogen) was mixed with 6 µl Cellfectin (Invitrogen) and incubated for 30 min at 22 °C. Two millilitres of Sf9 cells at mid-log phase (1 × 10⁶ cells ml⁻¹) were infected and incubated at 28 °C for 96 h. The supernatant was collected by centrifugation at 2,000 r.p.m. for 5 min in a clinical centrifuge. This viral sample was diluted to 1/100 and infected into 50 ml Sf9 cells at mid-log phase (1 × 10⁶). Virus was collected 96 h post-infection by centrifugation at 2,000 r.p.m. for 5 min. After amplifying the virus in Sf9 cells, H6-TIR1-Myc protein was expressed in High Five cells. One millilitre of High Five cells at the mid-log phase (1 × 10⁶) were infected with 1/500 dilution of the reamplified virus and cells were collected by centrifugation 24 h post-infection. Cells were washed once with the extraction buffer (described above) and then resuspended in 1 ml of the extraction buffer. Crude protein was isolated by sonication of the cells and the extract was cleared by centrifugation at 1,300 r.p.m. for 10 min at 4 °C. Total protein in the extract was estimated by the Bradford method. Approximately 100 µg total protein was used in the 250-µl pull-down reaction.

Received 10 January; accepted 11 March 2005.

- Davies, P. J. in *Plant Hormones Physiology, Biochemistry and Molecular Biology* (ed. Davies, P. J.) 1–12 (Kluwer Academic, Dordrecht, 1995).
- Went, F. W. Wushstoff und wachstum. *Rec. Trav. Bot. Neerl.* **25**, 1–116 (1928).
- Weijers, D. & Jurgens, G. Auxin and embryo axis formation: the ends in sight? *Curr. Opin. Plant Biol.* **8**, 32–37 (2005).
- Willemsen, V. & Scheres, B. Mechanisms of pattern formation in plant embryogenesis. *Annu. Rev. Genet.* **38**, 587–614 (2004).
- Napier, R. M., David, K. M. & Perrot-Rechenmann, C. A short history of auxin-binding proteins. *Plant Mol. Biol.* **49**, 339–348 (2002).
- Hagen, G. & Guilfoyle, T. Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol. Biol.* **49**, 373–385 (2002).
- Reed, J. W. Roles and activities of Aux/IAA proteins in *Arabidopsis*. *Trends Plant Sci.* **6**, 420–425 (2001).
- Liscum, E. & Reed, J. W. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol. Biol.* **49**, 387–400 (2002).
- Zenser, N., Ellsmore, A., Leasure, C. & Callis, J. Auxin modulates the degradation rate of Aux/IAA proteins. *Proc. Natl Acad. Sci. USA* **98**, 11795–11800 (2001).
- Gray, W. M., Kepinski, S., Rouse, D., Leyser, O. & Estelle, M. Auxin regulates SCF^{TIR1}-dependent degradation of Aux/IAA proteins. *Nature* **414**, 271–276 (2001).
- Tiwari, S. B., Wang, X. J., Hagen, G. & Guilfoyle, T. J. Aux/IAA proteins are active repressors, and their stability and activity are modulated by auxin. *Plant Cell* **13**, 2809–2822 (2001).
- Tian, Q., Nagpal, P. & Reed, J. W. Regulation of *Arabidopsis* SHY2/IAA3 protein turnover. *Plant J.* **36**, 643–651 (2003).
- Ramos, J. A., Zenser, N., Leyser, H. M. & Callis, J. Rapid degradation of Aux/IAA proteins requires conserved amino acids of domain II and is proteasome-dependent. *Plant Cell* **13**, 2349–2360 (2001).
- Dharmasiri, N., Dharmasiri, S., Jones, A. M. & Estelle, M. Auxin action in a cell-free system. *Curr. Biol.* **13**, 1418–1422 (2003).
- Kepinski, S. & Leyser, O. Auxin-induced SCF^{TIR1}-Aux/IAA interaction involves stable modification of the SCF^{TIR1} complex. *Proc. Natl Acad. Sci. USA* **101**, 12381–12386 (2004).
- Ouellet, F., Overvoorde, P. J. & Theologis, A. IAA17/AXR3. Biochemical insight into an auxin mutant phenotype. *Plant Cell* **13**, 829–842 (2001).
- Deshaies, R. J. SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu. Rev. Cell Dev. Biol.* **15**, 435–467 (1999).
- Jaakkola, P. et al. Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* **292**, 468–472 (2001).
- Cardozo, T. & Pagano, M. The SCF ubiquitin ligase: insights into a molecular machine. *Nature Rev. Mol. Cell Biol.* **5**, 739–751 (2004).
- Yoshida, Y. et al. E3 ubiquitin ligase that recognizes sugar chains. *Nature* **418**, 438–442 (2002).
- Yang, X. et al. The IAA1 protein is encoded by AXR5 and is a substrate of SCF^{TIR1}. *Plant J.* **40**, 772–782 (2004).
- Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M. & Vierstra, R. D. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11519–11524 (2002).
- Gray, W. M. et al. Identification of an SCF ubiquitin-ligase complex required for auxin response in *Arabidopsis thaliana*. *Genes Dev.* **13**, 1678–1691 (1999).
- Xie, D. X., Feys, B. F., James, S., Nieto-Rostro, M. & Turner, J. G. COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280**, 1091–1094 (1998).
- Kinoshita, T. et al. Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167–171 (2005).
- Frias, I. et al. A major isoform of the maize plasma membrane H(+)-ATPase: characterization and induction by auxin in coleoptiles. *Plant Cell* **8**, 1533–1544 (1996).
- Blatt, M. R. & Thiel, G. K⁺ channels of stomatal guard cells: bimodal control of the K⁺ inward-rectifier evoked by auxin. *Plant J.* **5**, 55–68 (1994).
- Philippart, K. et al. Auxin-induced K⁺ channel expression represents an essential step in coleoptile growth and gravitropism. *Proc. Natl Acad. Sci. USA* **96**, 12186–12191 (1999).
- Luthen, H., Claassen, M. & Bottger, M. Growth: progress in auxin research. *Prog. Bot.* **60**, 315–340 (1999).
- Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).

Acknowledgements We thank R. Dumas for technical assistance and C. Leblanc for help with insect cell culture. In addition, we thank S. Kepinski and O. Leyser for discussions. Research in the authors' laboratory is supported by grants from the NIH, the NSF and the DOE.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.E. (maestell@indiana.edu).

ARTICLES

The *Arabidopsis* F-box protein TIR1 is an auxin receptor

Stefan Kepinski^{1,2} & Ottoline Leyser¹

Despite 100 years of evidence showing a pivotal role for indole-3-acetic acid (IAA or auxin) in plant development, the mechanism of auxin perception has remained elusive. Central to auxin response are changes in gene expression, brought about by auxin-induced interaction between the Aux/IAA transcriptional repressor proteins and the ubiquitin-ligase complex SCF^{TIR1}, thus targeting for them proteolysis. Regulated SCF-mediated protein degradation is a widely occurring signal transduction mechanism. Target specificity is conferred by the F-box protein subunit of the SCF (TIR1 in the case of Aux/IAAs) and there are multiple F-box protein genes in all eukaryotic genomes examined so far. Although SCF-target interaction is usually regulated by signal-induced modification of the target, we have previously shown that auxin signalling involves the modification of SCF^{TIR1}. Here we show that this modification involves the direct binding of auxin to TIR1 and thus that TIR1 is an auxin receptor mediating transcriptional responses to auxin.

The small indolic compound indole-3-acetic acid (IAA or auxin) has a unique and seemingly pervasive position in the regulation of plant growth and development¹. An elaborate system of influx and efflux carriers regulates the distribution of auxin within plant tissues, and rapid auxin-induced changes in gene expression mediate various downstream responses^{2–4}. Despite more than a century of research underpinning these upstream and downstream events, the mechanism of auxin perception has remained elusive¹. The best-characterized auxin binding protein is the enigmatic, endoplasmic reticulum-localized, auxin binding protein 1 (ABP1)⁵. However, there is no evidence linking ABP1 to auxin-regulated gene expression.

Auxin induces transcription by targeting for degradation members of the Aux/IAA family of transcriptional repressor proteins, of which there are 29 in *Arabidopsis*^{6–10}. Aux/IAAs can dimerize with, and repress the activity of, transcriptional activators of the auxin response factor (ARF) family of DNA-binding proteins^{8,11}. Thus the degradation of Aux/IAAs leads to the derepression of ARF-mediated transcription^{8,10,12}. Aux/IAAs are targeted for degradation by ubiquitination, catalysed by an SCF-type ubiquitin-protein ligase^{6,13}. SCFs have a characteristic subunit structure, consisting of a SKP1 protein, a Cullin, an F-box protein, and RBX1 (refs 13, 14). SKP1 links the F-box protein to the Cullin, which in turn interacts with RBX1 (refs 13, 14). The RBX1–Cullin dimer catalyses the transfer of activated ubiquitin from a ubiquitin-conjugating enzyme to the target protein. Target specificity is conferred by the F-box protein, which includes an amino-terminal F-box motif mediating interaction with SKP1, and a carboxy-terminal protein–protein interaction domain thought to be involved in target selection^{13,14}. For *Arabidopsis* Aux/IAAs, the relevant F-box protein is the leucine-rich repeat TIR1 protein⁶, and probably a small family of close structural homologues¹⁵.

The interaction between Aux/IAAs and SCF^{TIR1} is central to auxin biology. Mutations in Aux/IAAs that severely reduce their interaction with SCF^{TIR1} also increase their stability and confer defects in auxin-induced gene expression, causing a wide range of auxin-related morphological phenotypes^{6,9,10,16}. A similar suite of effects result from mutations that disrupt SCF^{TIR1} function^{14,17–20}. Among these,

mutations in TIR1 have relatively mild phenotypes, presumably because of redundancy with its close homologues.

All of the Aux/IAA mutations that disrupt the interaction with SCF^{TIR1} map to a conserved region known as domain II^{6,9,10,16,21}. Indeed, a biotinylated peptide consisting of 17 amino acids from the core of domain II is sufficient to support auxin-regulated interaction with SCF^{TIR1} in pull-down assays²². In previously characterized examples, signal-regulated SCF–target interaction is mediated by signal-induced modification of the target, most commonly by phosphorylation¹³. However, this canonical mechanism is unlikely for Aux/IAAs because a wide range of pharmacological agents that disrupt classical signalling pathways do not affect the Aux/IAA–SCF^{TIR1} interaction^{9,23}. Furthermore, the Aux/IAA–SCF^{TIR1} interaction occurs in plant extracts from which membranes have been removed, indicating that the site of auxin perception in this response might be soluble²³. These results, and our previous observation that auxin influences the interaction by affecting SCF^{TIR1} rather than Aux/IAAs²² led us to speculate that the auxin signal transduction pathway upstream of Aux/IAA degradation is very short and/or unconventional in its configuration. Here we show that auxin promotes the interaction between SCF^{TIR1} and Aux/IAAs, and hence Aux/IAA degradation, by binding directly to TIR1.

An auxin receptor activity copurifies with TIR1

To explore further the mechanism by which auxin affects SCF^{TIR1}, we immunoprecipitated SCF^{TIR1} from extracts prepared from *tir1-1* mutant plants transgenically expressing Myc-tagged TIR1 protein (TIR1–Myc), using anti-Myc antibodies. We then resuspended the immunoprecipitate in extraction buffer and tested the ability of the proteins recovered in this way to support auxin-regulated interaction with the biotinylated Aux/IAA domain II peptide in a pull-down assay. The domain II peptide was recovered on streptavidin-coated agarose beads, and the amount of copurified TIR1–Myc was determined by western blotting. Addition of the natural auxin IAA resulted in a greatly enhanced copurification of TIR1–Myc with the Aux/IAA domain II peptide, in a dose-dependent manner, with clear activity when supplied at a concentration of 0.5 μ M (Fig. 1a, b). This shows that the proteins recovered in the immunoprecipitation

¹Department of Biology, University of York, Box 373, York YO10 5YW, UK. ²Umeå Plant Science Centre, SLU, S-90183, Umeå, Sweden.

include those sufficient to allow the auxin-enhanced SCF^{TIR1} -Aux/IAA interaction and hence the immunoprecipitate must include the auxin receptor for this response. The synthetic auxins 1-naphthaleneacetic acid (1-NAA) and 2,4-dichlorophenoxyacetic acid (2,4-D) were also able to promote the interaction, but with lower activity. 1-NAA has a clear promotive effect at $10\ \mu\text{M}$ (Fig. 1b), whereas 2,4-D is the least active auxin tested, with no clear effect on the SCF^{TIR1} -Aux/IAA interaction at concentrations less than $50\ \mu\text{M}$ (Fig. 1c). No effect was detected with the inactive auxin-like compound 2-NAA nor the weak acid benzoic acid (Fig. 1a, b). These dose-response profiles mirror those observed for similar pull-down assays performed on plant extracts^{6,23}, confirming that full auxin response capacity has been recovered in the immunoprecipitated proteins.

This short signal transduction pathway is likely to be rapid, and indeed the addition of $10\ \mu\text{M}$ 1-NAA to TIR1-Myc-containing plant extracts for only 5 min at 4°C was sufficient to enhance the recovery of TIR1-Myc in pull-down assays with the biotin-tagged Aux/IAA domain II peptide (Fig. 2a), and the recovery was saturated by 20 min after the addition of auxin to the assay (Fig. 2a). To determine whether the assayed interaction was dependent on the continued presence of auxin, we first recovered TIR1-Myc from plant extracts by using biotinylated Aux/IAA domain II peptide bound to streptavidin-agarose, as before. We then compared the amount of TIR1-Myc associated with the Aux/IAA domain II peptide after it had been washed in buffer either with or without added auxin. If auxin is continuously required to sustain the interaction, washing with active auxin (1-NAA) should preserve the interaction better than washing with inactive auxin (2-NAA). This was indeed found to be so (Fig. 2b).

That auxin is continuously required argues against its acting to promote a stable covalent modification and indicates that the direct binding of auxin to SCF^{TIR1} , or a tightly associated protein, might be sufficient to enhance SCF^{TIR1} -Aux/IAA binding. This mode of action for auxin is consistent with our previous work on the SCF^{TIR1} -Aux/IAA interaction²², in which we found that when SCF^{TIR1} was pretreated with auxin in plant extracts, immunoprecipitated, washed, and returned to plant extracts with an Aux/IAA domain II

target but no auxin, it was able to interact with the target significantly better than if no pretreatment with auxin was done. However, this interaction was very much less effective than if auxin was added afresh at the end of the experiment. Presumably auxin binds to the SCF during the pretreatment, and the washing procedure removes some but not all of the SCF-associated auxin, allowing some but not all of the enhanced Aux/IAA interaction to be maintained.

SCF^{TIR1} binds auxin directly

To test this possibility, we investigated whether auxin bound directly to the SCF^{TIR1} complex. We added radiolabelled auxin to pull-down assays from TIR1-Myc-containing plant extracts and compared the recovery of radiolabel in the presence of various concentrations of competing, unlabelled auxin. Unlabelled IAA was highly effective at reducing the recovery of tritiated IAA ($[^3\text{H}]\text{IAA}$) in the pull-down assay (Fig. 3a), showing that the interaction involves direct auxin binding. Fitting these data on the assumption of a single class of binding site indicates a dissociation constant, K_d , for IAA of about $25\ \text{nM}$. However, because this estimate of K_d is particularly sensitive to small differences in the background level of labelled auxin, the K_d for IAA should more cautiously be considered to be within the range $20\text{--}80\ \text{nM}$. Furthermore, the value does not take into account the unknown level of auxin and other auxin-binding sites in the plant extracts used for this assay. None the less, the data show that significant binding occurs at physiologically relevant concentrations of auxin. The synthetic auxins 2,4-D and 1-NAA were also able to compete with $[^3\text{H}]\text{IAA}$ for binding to the complex, but with lower activity than IAA (Fig. 3b): 1-NAA was about one order of magnitude, and 2,4-D, two orders of magnitude weaker in binding than IAA. The ability of each auxin to compete with IAA quantitatively mirrors the ability of the auxins to enhance the SCF^{TIR1} -Aux/IAA interaction (compare Fig. 3b with Fig. 1), further supporting the physiological relevance of the auxin binding.

No evidence of specific $[^3\text{H}]\text{IAA}$ binding was observed in control assays performed without seedling extracts (data not shown), indicating that auxin binds to the SCF^{TIR1} complex rather than to the Aux/IAA. This idea is further supported by our observation that increasing the level of TIR1-Myc in otherwise identical *tir1-1* mutant seedling extracts correlated with an increased auxin binding capacity in pull-down assays (Fig. 3c). In these experiments, $[^3\text{H}]\text{IAA}$ pull-down assays were performed with either *tir1-1* extract (0% TIR1-Myc), two-thirds *tir1-1* extract, one-third *tir1-1*/[TIR1-Myc] extract (33% TIR1-Myc) or *tir1-1*/[TIR1-Myc] (100% TIR1-Myc). Subtracting the radioactive labelling for 0% TIR1-Myc from the assays of extracts containing TIR1-Myc shows that the recovery of $[^3\text{H}]\text{IAA}$ in

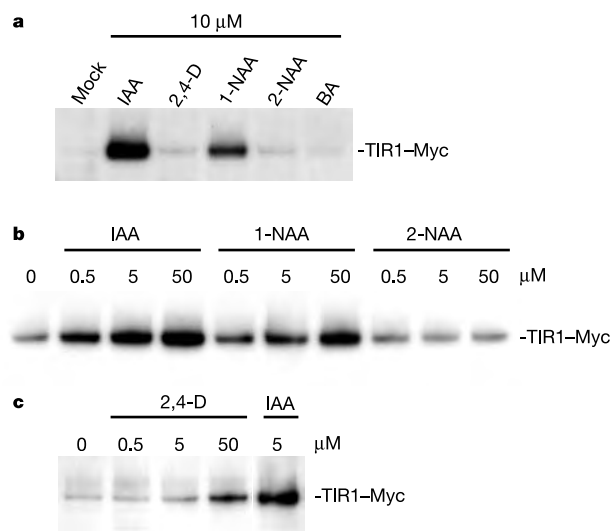


Figure 1 | Immunopurified TIR1-Myc retains the ability to interact with Aux/IAA domain II peptides in an auxin-dependent manner. **a**, Domain II peptide pull-down assays with immunopurified TIR1-Myc in the presence of $10\ \mu\text{M}$ IAA, 2,4-D, 1-NAA, 2-NAA and benzoic acid (BA) as indicated. The recovery of TIR1-Myc was assessed by immunoblotting with anti-c-Myc antibody. **b**, **c**, As in **a** but showing dose responses for IAA, 1-NAA and 2-NAA (**b**), and 2,4-D (**c**) over the concentration range $0\text{--}50\ \mu\text{M}$ as indicated.

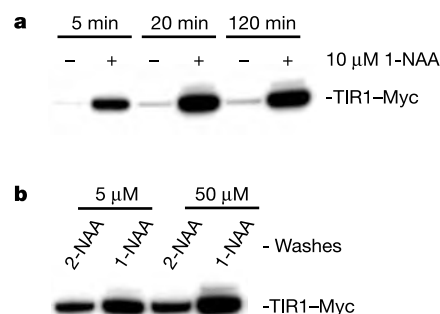


Figure 2 | The association and dissociation of TIR1-Aux/IAA complexes are rapid and indicates direct auxin binding in the complex. **a**, Anti-c-Myc immunoblot of domain II peptide pull-down assays with extracts prepared from *tir1-1*/[TIR1-Myc] seedlings. The assays were treated with and without $10\ \mu\text{M}$ 1-NAA and allowed to proceed for 5, 20 and 120 min as indicated. **b**, Otherwise identical pull-down assays with $10\ \mu\text{M}$ 1-NAA-treated *tir1-1*/[TIR1-Myc] seedling extract washed five times with either 2-NAA or 1-NAA at the concentrations indicated.

33% TIR1–Myc assays is 34.5% of that in 100% TIR1–Myc assays. This indicates quantitative [^3H]IAA binding with respect to TIR1–Myc dose and strongly indicates that SCF^{TIR1} binds auxin. We suspect that the auxin-binding activity apparent in [^3H]IAA pull-downs from 0% TIR1–Myc extracts is due to the presence of functional homologues of TIR1 in the *Arabidopsis* genome. This is supported by the fact that extracts of quadruple mutants lacking TIR1 and three closely related family members no longer possess a specific auxin-binding activity in pull-down assays¹⁵.

TIR1 is an auxin receptor

These results show that auxin binds directly to the SCF^{TIR1} complex. The only known auxin-specific subunit of the complex is TIR1. We therefore tested whether expressing TIR1 in a heterologous, non-auxin-responsive system was sufficient for an auxin-regulated interaction with Aux/IAA targets. To do this we used a *Xenopus laevis* embryo expression system²⁴. Two-cell *Xenopus* embryos were bilaterally injected with capped TIR1–Myc mRNA. The embryos were harvested at stage 13 and extracts were made for use in the pull-down assay. *Xenopus*-expressed TIR1–Myc was found to interact with Aux/IAA domain II peptide in an auxin-regulated and dose-dependent way (Fig. 4a, b). As with extracts from plant tissue, washing the pulled-down protein complexes with 1-NAA but not 2-NAA was found to increase the amount of TIR1–Myc associated with the target peptide (Fig. 4c). In combination with our previous observations²² that auxin modifies SCF^{TIR1}, this result indicates that TIR1 is the auxin receptor for this response.

To test this idea further, we repeated the [^3H]IAA-binding experiments with the use of *Xenopus* extracts. To eliminate the possibility that *Xenopus* extracts possess an auxin-binding activity in this assay, we performed [^3H]IAA pull-downs with uninjected *Xenopus* extracts. No specific [^3H]IAA binding was detected (Fig. 4d). However, the same assay performed with extracts of embryos injected with TIR1–Myc mRNA showed clear evidence of saturable binding of auxin as judged by the ability of unlabelled IAA to displace [^3H]IAA (Fig. 4e). Although not directly comparable, these data are similar to those for Fig. 3b, indicating that heterologously expressed and *Arabidopsis*-expressed TIR1 proteins might possess similar IAA binding characteristics.

Auxin-induced TIR1–Aux/IAA binding requires the TIR1 F-box

These data show that addition of only two plant proteins to *Xenopus* extracts, namely TIR1–Myc and the domain II peptide, confers saturable auxin-binding activity and full competence for auxin-stimulated interaction between the two proteins. The Aux/IAA proteins, including domain II, are specific to plants; they have no homologues in animals, nor is there any evidence that IAA is active as a signal in animal systems. Thus it is highly unlikely that *Xenopus* has an auxin receptor capable of associating with, and brokering the interaction between, the domain II peptide and TIR1. None the less, it is possible that proteins conserved between *Xenopus* and plants participate in a receptor complex. In particular, we speculated whether TIR1 interacts with endogenous SCF components in *Xenopus*, and whether this association is required for the auxin-regulated interaction with the domain II peptide.

F-box proteins are brought into SCFs by the interaction of the F-box motif with SKP1. In *Arabidopsis*, TIR1 interacts with the SKP1-like protein ASK1. ASK1 shares significant homology with the *X. laevis* Skp1 protein, particularly in the C-terminal helical domains proposed to form the major part of the binding surface with the F-box motif²⁵. Because the F-box of TIR1 could reasonably be expected to interact with *Xenopus* Skp1, we tested whether the F-box was required for the auxin-stimulated TIR1–domain II peptide interaction. We expressed F-box-deleted TIR1–Myc, lacking the first 47 amino acids of the protein ($\Delta\text{F-TIR1-Myc}$), in *Xenopus* embryos and compared the recovery of this mutant protein with that of full-length TIR1–Myc in domain II peptide pull-down assays from cell extracts. The results show that deletion of the F-box abolishes auxin-stimulated interaction between TIR1–Myc and the domain II peptide (Fig. 5a).

We next tested whether Skp1 copurifies with TIR1–Myc in *Xenopus* extracts. To do this, TIR1–Myc was expressed in *Xenopus* as described above; western blots of proteins pulled down with the domain II peptide in the presence and absence of auxin were probed with anti-Skp1 antibody. Despite the abundance of Skp1 detected in the *Xenopus* extracts, little or no Skp1 copurified with TIR1 in these experiments (Fig. 5b). Therefore, unless the Skp1 antibody has a particularly low affinity and Skp1 is an extremely abundant protein in *Xenopus* embryos, most of the TIR1–Myc recovered in these

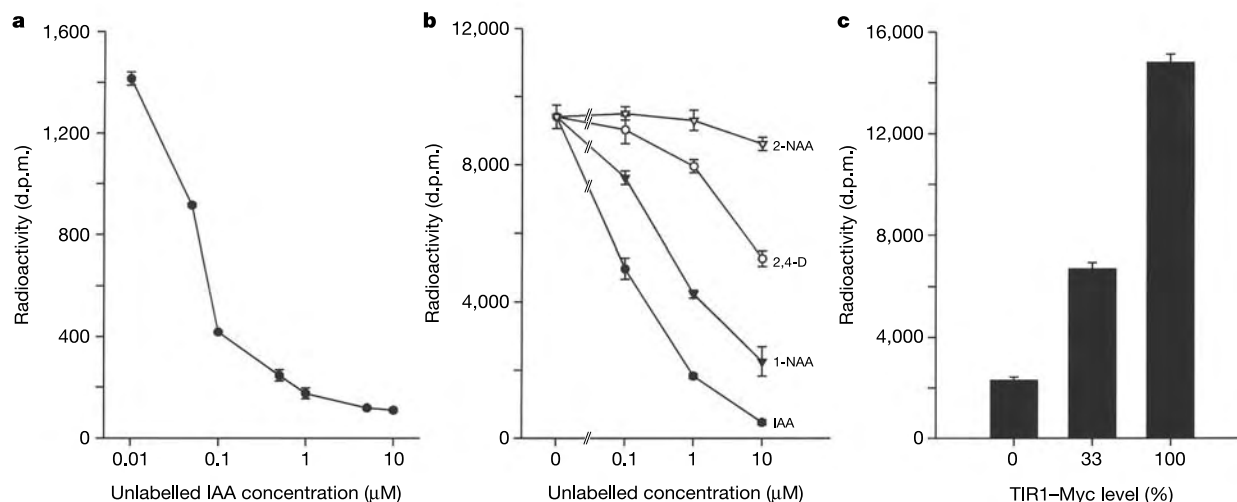


Figure 3 | The TIR1–Aux/IAA interaction involves direct auxin binding. **a**, GST–AXR2 pull-down assays from *tir1-1*[TIR1–Myc] seedling extract with 0.01 μM [^3H]IAA and increasing concentrations of unlabelled IAA. Binding of [^3H]IAA was assessed by scintillation counting and is expressed as disintegrations per minute (d.p.m.). **b**, As in **a**, except that the experiment was performed with the [^3H]IAA concentration held constant at 0.1 μM and

with increasing concentrations of unlabelled IAA (filled circles), 2,4-D (open circles), 1-NAA (filled triangles) or 2-NAA (open triangles) as indicated. **c**, GST–AXR2 pull-down assays from otherwise identical *tir1-1* seedling extracts containing 0.1 μM [^3H]IAA and increasing levels of TIR1–Myc as indicated. Error bars indicate s.e.m. ($n = 3$).

experiments is not involved in interaction with *Xenopus* Skp1. This indicates that the incorporation of TIR1 into an SCF might not be required for auxin binding or Aux/IAA interaction. Therefore, despite the requirement for the F-box, there is apparently no requirement for SKP1 interaction. These findings raise the interesting possibility that the F-box of TIR1 could have another function in addition to mediating SKP1 interaction. Alternatively, it is possible that mutation or deletion of the F-box simply compromises TIR1 folding in a non-specific way.

To investigate less drastic changes to the TIR1 F-box, we tested the effect of point mutations at three conserved positions (Pro 10 → Ala (P10A), Val 33 → Ala (V33A) and Lys 35 → Ala (K35A)) in the F-box on auxin-stimulated interaction with the domain II peptide in wheatgerm extracts. All the TIR1–Myc derivatives were efficiently expressed by coupled *in vitro* transcription/translation in wheatgerm extract (Fig. 5d). In pull-down assays from these extracts, wild-type TIR1–Myc supported auxin-stimulated interaction with the domain II peptide (Fig. 5c). In comparison with wild type, the P10A mutation severely reduced, and the V33A and K35A mutations abolished, the auxin-enhanced recovery of the F-box mutant variants (Fig. 5c). Although a requirement for SKP1 binding cannot definitely be excluded, these results provide some support for the idea that the F-box of TIR1 has a specific function in addition to SKP1 binding, such as auxin or Aux/IAA binding.

Discussion

We had previously reported that auxin acts to promote SCF^{TIR1}–Aux/IAA interaction, and hence Aux/IAA degradation, by modifying SCF^{TIR1} rather than the Aux/IAAs²². Here we have shown that this

modification is the direct binding of auxin to TIR1 and thus that TIR1 is an auxin receptor that mediates transcriptional responses to auxin. Although auxin perception by TIR1 and its close homologues is sufficient to account for a substantial part of the auxin response, there is good evidence for an additional extracellular site of auxin perception. For example, changes in ion transport associated with the early stages of auxin-induced growth can be inhibited by extracellular treatment with anti-ABP1 antibodies, indicating that ABP1 might act as an extracellular auxin receptor for this response²⁶.

As well as providing key evidence of an auxin receptor function for TIR1, our work with heterologously expressed TIR1 also answers the previously open question of whether TIR1 binds Aux/IAA targets directly or indirectly through some kind of adaptor protein, indicating that binding is direct and that in all probability no additional components are required. If this is indeed so, we further conclude that SCF^{TIR1} and associated protein degradation machinery, the Aux/IAAs, and their binding partners the ARFs represent a complete signal transduction cascade from auxin to gene expression.

Precisely how the binding of auxin to TIR1 promotes TIR1–Aux/IAA interaction is not known. The simplest mechanism would be one in which the binding of auxin to TIR1 causes a conformational change that allows Aux/IAA binding. Alternatively, it is possible that TIR1 is able to interact with Aux/IAAs in the absence of auxin but that the association is significantly stabilized by subsequent auxin binding. It is also not clear where and how many molecules of IAA bind to TIR1. These details will be the subject of future studies.

Our results also describe a novel mode of regulation of SCF–target interaction in which the signal that prompts the interaction binds directly to the F-box protein; our observations therefore define a new

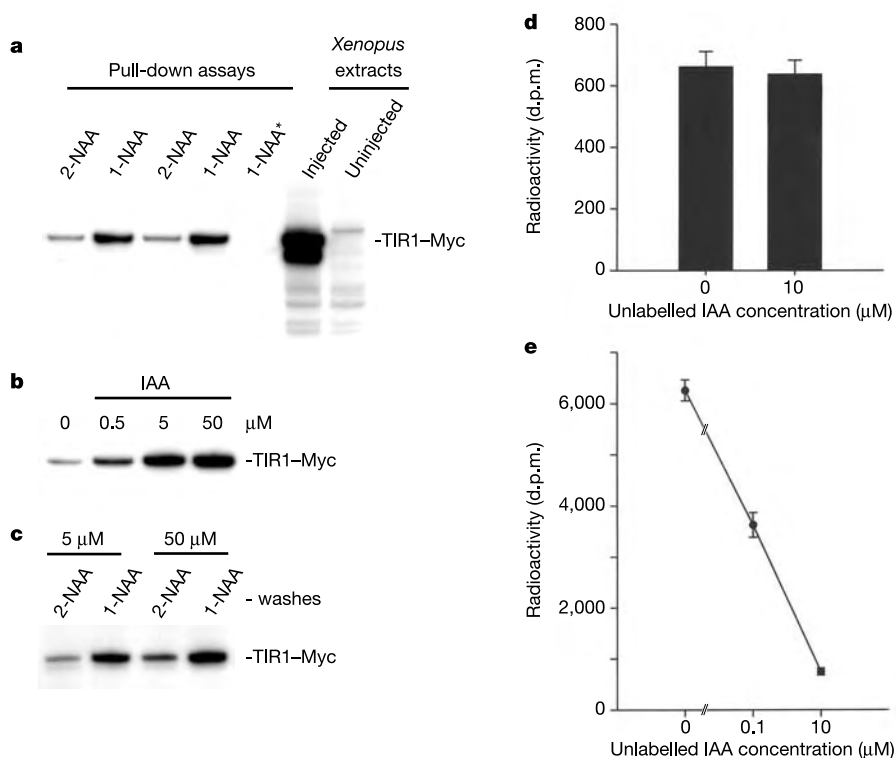


Figure 4 | TIR1-Myc expressed in *X. laevis* embryos supports auxin-enhanced interaction with Aux/IAA domain II peptides and direct auxin binding. **a**, Anti-c-Myc immunoblot. Left five lanes, domain II peptide pull-down assays, in the presence of 10 μ M 1-NAA or 2-NAA as indicated, with extracts prepared from *Xenopus* embryos expressing TIR1-Myc. The lane marked with an asterisk is a control pull-down assay with an extract of uninjected embryos. Right two lanes, crude extracts of *Xenopus* embryos either uninjected or injected with 2 ng of capped TIR1-Myc mRNA as indicated. **b**, As in **a**, but showing IAA dose-response pull-down assays over

the range 0–50 μ M as indicated. **c**, Otherwise identical pull-down assays, washed five times with either 2-NAA or 1-NAA at the concentrations indicated, with extracts prepared from *Xenopus* embryos expressing TIR1-Myc. **d**, Pull-down assays, in the presence of 0.1 μ M [³H]IAA and either with or without 10 μ M unlabelled IAA as indicated, with extracts prepared from uninjected *Xenopus* embryos. **e**, Pull-down assays, in the presence of 0.1 μ M [³H]IAA and increasing concentrations of unlabelled IAA as indicated, with extracts prepared from injected *Xenopus* embryos expressing TIR1-Myc. Error bars indicate s.e.m. ($n = 3$).

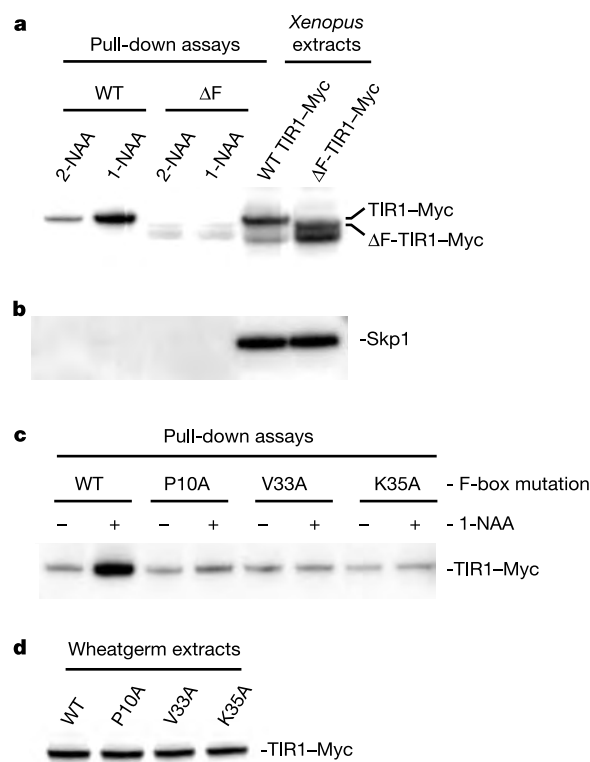


Figure 5 | The F-box motif of TIR1 is required for auxin-induced TIR1-Aux/IAA interaction. **a**, Anti-c-Myc immunoblot. Left four lanes, domain II peptide pull-down assays with extracts prepared from *Xenopus* embryos expressing full-length TIR1-Myc (WT) or F-box-deleted TIR1-Myc (ΔF), in the presence of 10 μM 1-NAA or 2-NAA as indicated. Right two lanes, crude extracts of *Xenopus* embryos injected with 2 ng of capped WT-TIR1-Myc or ΔF -TIR1-Myc mRNA as indicated. **b**, As in **a**, but showing the detection of *Xenopus* Skp1 by anti-Skp1 immunoblot. **c**, Anti-c-Myc immunoblot of domain II peptide pull-down assays with wheat germ extracts expressing wild-type TIR1 (WT) and the TIR1 F-box mutant variants P10A, V33A and K35A, in the presence or absence of 10 μM 1-NAA as indicated. **d**, Anti-c-Myc immunoblot of wheat germ extracts expressing wild-type and F-box mutant variants of TIR1.

class of receptor. Plants seem exceptional among eukaryotes in the extent to which SCF-mediated proteolysis is used. The *Arabidopsis* genome, for example, encodes more than 700 F-box proteins^{27,28}, many more than are found in other similarly complex organisms (for example there are 74 in humans and 53 in mice^{29,30}). Plant development depends on the continuous integration of numerous signals, many of them small molecules. With so many uncharacterized SCF complexes in plants, and indeed other eukaryotes, it will be interesting to determine whether SCF-based mechanisms of signal perception are widespread.

METHODS

Plant materials, peptides and recombinant proteins. The lines *tir1-1* and *tir1-1*[TIR1-Myc], growth conditions, and methods of TIR1-Myc induction and seedling extraction have been described previously^{6,17}. The synthetic domain II peptide and the fusion protein glutathione S-transferase (GST)-AXR2 have also been described previously^{6,22}.

***Xenopus* expression of TIR1-Myc.** The TIR1-Myc coding sequence was amplified from pGB31 by polymerase chain reaction (PCR)¹⁷ by using the oligonucleotides 5'-TTGGATCCATGCAGAAGCGAATAGCCTTG-3' and 5'-ATGAATTCCTTATAATCCGTTAGTAGTAATGATTG-3', and cloned as a *Bam*HI-*Eco*RI fragment into the expression vector pCS2 + downstream of the SP6 promoter to create the plasmid pXL TIR1-Myc. To create the plasmid pXL ΔF -TIR1-Myc, the coding sequence for all except the first 47 amino acids of TIR1-Myc was PCR-amplified with the oligonucleotides 5'-TTGGAT

CCATGGTCTTCATCGGGAAGT-3' and 5'-ATGAATTCCTTATAATCCGTTAGTAGTAATGATTG-3' and cloned as above. Linearized pXL TIR1-Myc and pXL ΔF -TIR1-Myc were used to synthesize capped TIR1-Myc mRNA and capped ΔF -TIR1-Myc mRNA, respectively, with the SP6 Megascript transcription kit (Ambion) with a modified protocol in which a 1:10 ratio of GTP to 5 mM m⁷G(5')Gppp(5')G cap was used. After synthesis, RNAs were subjected to sequential precipitation with propan-2-ol and 2.5 M LiCl. *X. laevis* embryos were obtained by *in vitro* fertilization, dejellied and cultured in 0.1 $\times$ normal amphibian medium (NAM) at 23 °C (ref. 24). Injections were performed on two-cell embryos (stage 2 (ref. 31)) in 0.3 $\times$ NAM plus 5% Ficoll type-400. TIR1-Myc mRNA (1 ng) in 10 nl was injected bilaterally (total 2 ng) and the injected embryos were incubated in 0.1 $\times$ NAM at 22 °C until stage 13 (ref. 31). Uninjected stage-13 control embryos for each set of injections were also collected. Crude extracts of embryos were made by lysing 100 embryos in 1,300 μl of lysis buffer (137 mM NaCl, 20 mM Tris-HCl pH 7.5, 2 mM EDTA, 1% Nonidet P40 containing 10 mM β -glycerophosphate, 1 mM phenylmethylsulphonyl fluoride, 1 mM dithiothreitol, 1 mM NaF, 10 μM MG132 and 2 μM pepstatin A). Extracts were cleared by centrifugation at 13,000g at 4 °C for 20 min. Extracts were used immediately.

***In vitro* transcription/translation of TIR1-Myc.** The TIR1-Myc coding sequence was PCR-amplified from pGB31 (ref. 17) by using the oligonucleotides 5'-TTGGATCCATGCAGAAGCGAATAGCCTTG-3' and 5'-ATGAATTCCTTATAATCCGTTAGTAGTAATGATTG-3' and cloned as a *Bam*HI-*Eco*RI fragment into the expression vector pcDNA3.1/V5-HisB (Invitrogen) downstream of the T7 promoter (and maintaining the original stop codon) to create the plasmid pWG TIR1-Myc. F-box mutant variants were created with the Quikchange site-directed mutagenesis kit (Stratagene) and complementary oligonucleotides with the sequences 5'-CGAATAGCCTTGTCGTTGTCAGAAGAGGTACTAGAGCATGTG-3' (pWG TIR1-MycP10A), 5'-GGAAGTCACTCTCTCTGGCGTGAAGTCATGGTACGAGATCG-3' (pWG TIR1-MycV33A) and 5'-CAGTCTCTCTGGTGTGCGCGTCATGGTACGAGATCGAGCGGT-3' (pWG TIR1-MycK35A). Linearized preparations of these plasmids were used with the TNT Coupled Wheat Germ Extract System (Promega) to synthesize wild-type and variant TIR1-Myc by *in vitro* transcription/translation (IVTT) in accordance with the manufacturer's instructions.

Immunoprecipitations and pull-down assays. TIR1-Myc was immunopurified with a 500- μl bed volume of anti-c-Myc 9E10-Sepharose (Covance) against 35 mg of crude *tir1-1*[TIR1-Myc] extract. The immunoprecipitates were washed three times in extraction buffer (EB; 0.15 M NaCl, 0.5% Nonidet P40, 0.1 M Tris-HCl pH 7.5, containing 10 mM β -glycerophosphate, 1 mM phenylmethylsulphonyl fluoride, 1 mM dithiothreitol, 1 mM NaF, 10 μM MG132 and 2 μM pepstatin A) and eluted with an equal bed volume of EB containing 400 $\mu g ml^{-1}$ Myc peptide (Sigma). Pull-down assays with immunopurified TIR1-Myc were performed by combining 75 μl of eluted TIR1-Myc immunoprecipitate with 6.5 μg of biotinylated domain II peptide and 425 μl of EB containing 1 mg ml^{-1} BSA (Sigma) with auxin treatments as indicated. The assays were incubated for 30 min at 4 °C with mixing, recovered on streptavidin-agarose (Novagen) and washed three times in EB. Standard pull-down assays from plant extracts have been described previously²². Pull-down assays with *Xenopus*-expressed TIR1-Myc were performed by combining 75 μl of TIR1-Myc RNA-injected *Xenopus* extract with 6.5 μg of biotinylated domain II peptide and 925 μl of EB containing 1 mg ml^{-1} BSA with auxin treatments as indicated. The assays were incubated for 30 min at 4 °C and recovered on streptavidin-agarose as before. Pull-down assays with wheat germ-expressed TIR1-Myc were performed by combining 45 μl of IVTT reaction extract with 6.5 μg of biotinylated domain II peptide and 955 μl of EB containing 1 mg ml^{-1} BSA with auxin treatments as indicated. The assays were incubated for 30 min at 4 °C and recovered on streptavidin-agarose as before. For pull-down/auxin washing experiments, 10 μM 1-NAA-treated pull-downs were performed as described above except that the final washes consisted of five 3-min washes in EB containing either 5 μM or 50 μM 1-NAA or 2-NAA as indicated. The final processing of pull-down assays including electrophoresis, western transfer and blotting with anti-Myc antibodies has been described previously²². The immunodetection of *Xenopus* Skp1 was performed with a 1:500 dilution of polyclonal anti-Skp1 antibody (Abcam), followed by a 1:4000 dilution of goat anti-rabbit IgG-peroxidase conjugate (Abcam), followed by chemiluminescent detection with ECL plus reagents (Amersham).

Binding assays with tritiated IAA. [³H]IAA (3-[5(n)-³H]indolylacetic acid, specific radioactivity 23 Ci mmol⁻¹) was purchased from Amersham Biosciences. For [³H]IAA pull-down assays from plant extracts, 0.01 or 0.1 μM [³H]IAA was added to about 3 mg of crude seedling extract and 10 μg of GST-AXR2 (immobilized on glutathione-Sepharose beads) in a volume of 1 ml. Unlabelled IAA, 2,4-D, 1-NAA and 2-NAA were added at the concentrations indicated. *tir1-1*[TIR1-Myc] seedling extract was used except in the experiments with

varied TIR1–Myc levels, in which *tir1-1* and *tir1-1*[TIR1–Myc] extracts were diluted to equal total protein concentration and combined as indicated. For [³H]IAA pull-down assays from *Xenopus* extracts, 2,600 µl of extract of *Xenopus* embryos (either uninjected or injected with TIR1–Myc mRNA) was combined with 10 µg of GST–AXR2 and 0.1 µM [³H]IAA, either without or with unlabelled IAA at the concentrations indicated. All assays were incubated for 30 min at 4 °C with mixing and then washed twice briefly in 750 µl of ice-cold EB, carefully removing as much of each wash as possible. The washed beads were resuspended in 100 µl of water and added to 3 ml of Ultima Gold XR scintillation fluid (Perkin Elmer) and vortex-mixed briefly. After 1 h, scintillation counting was performed in a Packard TriCarb2900 Liquid Scintillation Analyser (Perkin Elmer). All data points are the mean of at least three replicates and each experiment was performed independently at least three times.

Received 10 January; accepted 11 March 2005.

- Leyser, O. Molecular genetics of auxin signaling. *Annu. Rev. Plant Biol.* **53**, 377–398 (2002).
- Berleth, T., Krogan, N. T. & Scarpella, E. Auxin signals—turning genes on and turning cells around. *Curr. Opin. Plant Biol.* **7**, 553–563 (2004).
- Friml, J. Auxin transport—shaping the plant. *Curr. Opin. Plant Biol.* **6**, 7–12 (2003).
- Abel, S. & Theologis, A. Early genes and auxin action. *Plant Physiol.* **111**, 9–17 (1996).
- Napier, R. M., David, K. M. & Perrot-Rechenmann, C. P. A short history of auxin-binding proteins. *Plant Mol. Biol.* **49**, 339–348 (2002).
- Gray, W. M., Kepinski, S., Rouse, D., Leyser, O. & Estelle, M. Auxin regulates SCF^{TIR1}-dependent degradation of AUX/IAA proteins. *Nature* **414**, 271–276 (2001).
- Zenser, N., Ellsmore, A., Leasure, C. & Callis, J. Auxin modulates the degradation rate of Aux/IAA proteins. *Proc. Natl Acad. Sci. USA* **98**, 11795–11800 (2001).
- Tiwari, S. B., Wang, X.-J., Hagen, G. & Guilfoyle, T. J. AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. *Plant Cell* **13**, 2809–2822 (2001).
- Tian, Q., Nagpal, P. & Reed, J. W. Regulation of *Arabidopsis* SHY2/IAA3 protein turnover. *Plant J.* **36**, 643–651 (2003).
- Liscum, E. & Reed, J. W. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol. Biol.* **49**, 387–400 (2002).
- Ulmasov, T., Hagen, G. & Guilfoyle, T. J. Activation and repression of transcription by auxin-response factors. *Proc. Natl Acad. Sci. USA* **11**, 5844–5849 (1999).
- Tiwari, S. B., Hagen, G. & Guilfoyle, T. J. Aux/IAA proteins contain a potent transcriptional repression domain. *Plant Cell* **16**, 533–543 (2004).
- Deshaies, R. J. SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu. Rev. Cell Dev. Biol.* **15**, 435–467 (1999).
- Moon, J., Parry, G. & Estelle, M. The ubiquitin–proteasome pathway and plant development. *Plant Cell* **16**, 3181–3195 (2004).
- Dharmasiri, N., Dharmasiri, S. & Estelle, M. The F-box protein TIR1 is an auxin receptor. *Nature* doi:10.1038/nature03543 (this issue).
- Yang, X. *et al.* The IAA1 protein is encoded by AXR5 and is a substrate of SCF^{TIR1}. *Plant J.* **40**, 772–782 (2004).
- Gray, W. M. *et al.* Identification of an SCF ubiquitin-ligase complex required for auxin response in *Arabidopsis thaliana*. *Genes Dev.* **13**, 1678–1691 (1999).
- Gray, W. M., Hellmann, H., Dharmasiri, S. & Estelle, M. Role of the *Arabidopsis* RING-H2 protein RBX1 in RUB modification and SCF function. *Plant Cell* **14**, 2137–2144 (2002).
- Ruegger, M. *et al.* The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast Grr1p. *Genes Dev.* **12**, 198–207 (1998).
- Lincoln, C., Britton, J. H. & Estelle, M. Growth and development of the *axr1* mutants of *Arabidopsis*. *Plant Cell* **2**, 1071–1080 (1990).
- Ramos, J. A., Zenser, N., Leyser, O. & Callis, J. Rapid degradation of auxin/indoleacetic acid proteins requires conserved amino acids of domain II and is proteasome dependent. *Plant Cell* **13**, 2349–2360 (2001).
- Kepinski, S. & Leyser, O. Auxin-induced SCF^{TIR1}–Aux/IAA interaction involves stable modification of the SCF^{TIR1} complex. *Proc. Natl Acad. Sci. USA* **101**, 12381–12386 (2004).
- Dharmasiri, N., Dharmasiri, S., Jones, A. M. & Estelle, M. Auxin action in a cell-free system. *Curr. Biol.* **13**, 1418–1422 (2003).
- Isaacs, H. V., Pownall, M. E. & Slack, J. M. Regulation of Hox gene expression and posterior development by the *Xenopus* caudal homologue *Xcad3*. *EMBO J.* **17**, 3413–3427 (1998).
- Schulman, B. A. *et al.* Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex. *Nature* **408**, 381–386 (2000).
- Leblanc, N. *et al.* A novel immunological approach establishes that the auxin-binding protein, Nt-abp1, is an element involved in auxin signaling at the plasma membrane. *J. Biol. Chem.* **274**, 28314–28320 (1999).
- Risseuw, E. P. *et al.* Protein interaction analysis of SCF ubiquitin E3 ligase subunits from *Arabidopsis*. *Plant J.* **34**, 753–767 (2003).
- Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M. & Vierstra, R. D. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11519–11524 (2004).
- Smalle, J. & Vierstra, R. D. The ubiquitin 26S proteasome proteolytic pathway. *Annu. Rev. Plant Biol.* **55**, 555–590 (2004).
- Semple, C. A. M. RIKEN GER Group, GSL Members. The comparative proteomics of ubiquitination in mouse. *Genome Res.* **13**, 1389–1394 (2003).
- Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis* (Daudin) 2nd edn (North-Holland, Amsterdam, 1967).

Acknowledgements We thank R. Napier for invaluable advice on experimental design, numerous helpful discussions and critical reading of the manuscript; P. J. Burks and H. Isaacs for technical assistance with *Xenopus* expression; J. Hoggett for help with data analysis; C. Kleanthous, W. Gray, M. Estelle and T. Sieberer for discussions; and G. Sandberg and S. Day for critical reading of the manuscript. This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.K. (ssk2@york.ac.uk) or O.L. (hmol1@york.ac.uk).

ARTICLES

A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity

Carola G. Vinuesa¹, Matthew C. Cook², Constanza Angelucci¹, Vicki Athanasopoulos¹, Lixin Rui¹, Kim M. Hill¹, Di Yu¹, Heather Domaschensch¹, Belinda Whittle^{1,3}, Teresa Lambe⁴, Ian S. Roberts⁵, Richard R. Copley⁶, John I. Bell⁴, Richard J. Cornall⁴ & Christopher C. Goodnow^{1,3}

Despite the sequencing of the human and mouse genomes, few genetic mechanisms for protecting against autoimmune disease are currently known. Here we systematically screen the mouse genome for autoimmune regulators to isolate a mouse strain, *sanroque*, with severe autoimmune disease resulting from a single recessive defect in a previously unknown mechanism for repressing antibody responses to self. The *sanroque* mutation acts within mature T cells to cause formation of excessive numbers of follicular helper T cells and germinal centres. The mutation disrupts a repressor of ICOS, an essential co-stimulatory receptor for follicular T cells, and results in excessive production of the cytokine interleukin-21. *sanroque* mice fail to repress diabetes-causing T cells, and develop high titres of autoantibodies and a pattern of pathology consistent with lupus. The causative mutation is in a gene of previously unknown function, *roquin* (*Rc3h1*), which encodes a highly conserved member of the RING-type ubiquitin ligase protein family. The Roquin protein is distinguished by the presence of a CCCH zinc-finger found in RNA-binding proteins, and localization to cytosolic RNA granules implicated in regulating messenger RNA translation and stability.

Autoimmune diseases such as systemic lupus erythematosus (SLE) or type 1 diabetes result from failures of the immune system to repress self-reactive T cell responses and the formation of autoantibodies. Susceptibility to SLE occurs in clusters with a strong heritable component, but the genetic basis of this complex cellular response has only begun to be illuminated¹.

During normal immune responses, T lymphocytes recognizing a foreign antigen help antigen-specific B lymphocytes to make antibody through the pairing of T-cell and B-cell surface ligands and receptors, notably CD40 ligand (CD40L)–CD40, CD28–B7.1/B7.2 and ICOS–ICOS ligand (ICOSL), and by secretion of the cytokines interleukin (IL)-4, IL-5, IL-21 and interferon (IFN)- γ ^{2–10}. After receiving help from T helper type 1 (T_H1) or T_H2 cells in the T-cell-rich zones of secondary lymphoid organs, B cells undergo differentiation either outside follicles into short-lived plasma cells or, through germinal centres within follicles, into high-affinity, somatically mutated, long-lived plasma cells and memory cells¹¹. Most autoantibodies bind self-antigen with high affinity, indicating that they have differentiated through the latter pathway. However, current explanations for the repression of autoantibody responses focus on clonal deletion, anergy and the regulation of self-reactive T and B cells during the differentiation of immature lymphocytes or during subsequent interactions between dendritic cells, T cells and B cells in the T-cell-rich zones. Little is known about the processes for repressing self-reactive cells during the critical follicular/germinal centre stage.

A distinct subset of T cells, follicular helper T (T_{FH}) cells, selects mutated high-affinity B cells within germinal centres¹². T_{FH} cells are emerging as a cellular subset with a functional programme different

from that of extrafollicular T_H1 or T_H2 T cells: they contain high levels of ICOS^{13,14} and have distinct patterns of gene expression of cytokines (predominantly IL-21), chemokine receptors and transcription factors^{15–17}. Mutations in ICOS lead to profound defects in T_{FH} cell function for germinal centre antibody responses^{3–7,18}. Here we describe the discovery of a new ubiquitin ligase family member, Roquin, that is an essential negative regulator of T_{FH} cells and autoantibody responses.

Systematic screen for autoimmune regulators

We developed a systematic strategy to identify mechanisms for repressing anti-self immune responses by treating male C57BL/6 mice with ethylnitrosourea to generate single-base substitutions at an estimated rate of 1 per 0.5 megabases (Mb) in the germline, breeding the variant genome sequences to homozygosity, and screening for autoimmunity with the standard clinical test for antinuclear autoantibodies (ANAs). The first pedigree identified was named *sanroque* because of accompanying lymphadenopathy (Fig. 1a, b). *sanroque*-homozygous (*san/san*) females develop ANAs by 6–7 weeks of age and males by 8–16 weeks (Supplementary Fig. 1a). Additional testing revealed that *sanroque* mice had the following features typical of SLE (Fig. 1a–f, i; Supplementary Fig. 1b): high-affinity antibodies against double-stranded DNA (dsDNA), focal proliferative glomerulonephritis with deposition of IgG-containing immune complexes, necrotizing hepatitis, anaemia, and autoimmune thrombocytopenia in CBA *sanroque* mice (Fig. 1j, i).

T-cell autonomous defect

By 7 weeks of age, all *san/san* mice develop an enlarged spleen and

¹Division of Immunology and Genetics, John Curtin School of Medical Research, The Australian National University, ²The Australian National University Medical School, and

³Australian Phenomics Facility, Canberra, ACT 2601, Australia. ⁴Henry Wellcome Building of Molecular Physiology, University of Oxford, Oxford OX3 9DU, UK. ⁵Department of Cellular Pathology, John Radcliffe Hospital, Oxford OX3 9DU, UK. ⁶Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.

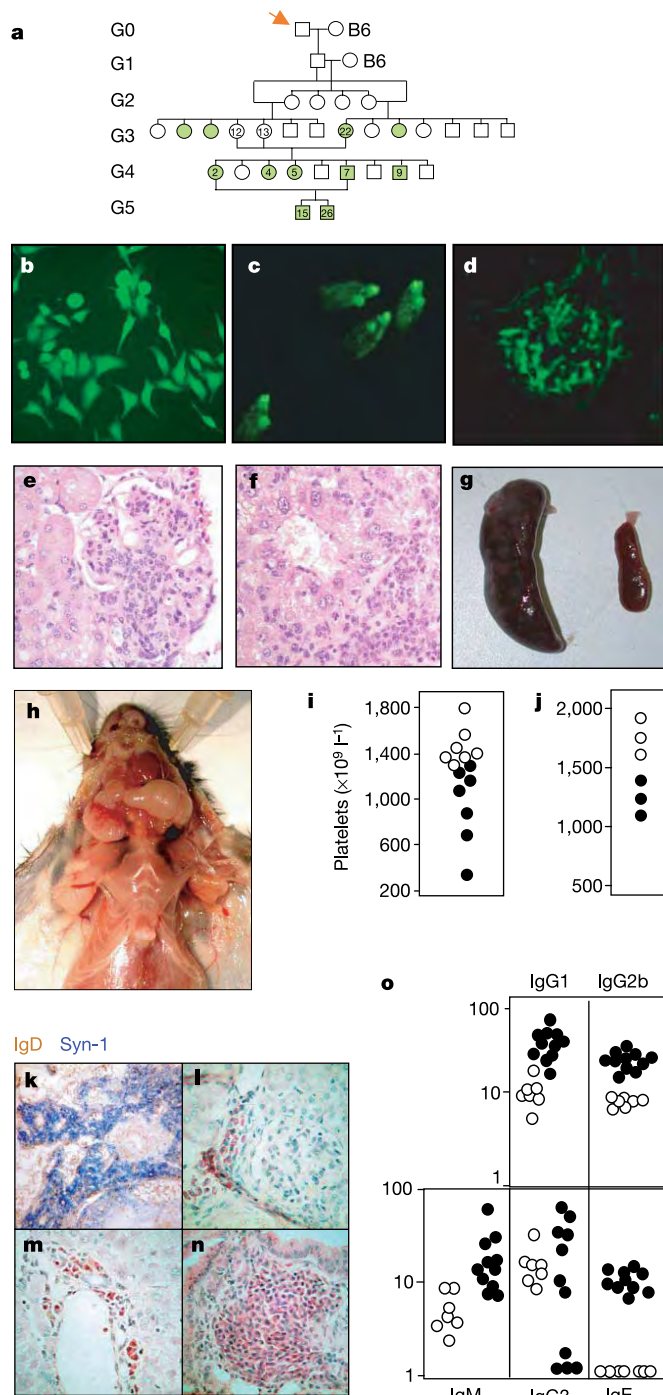


Figure 1 | Lupus-like pathology in *sanroque* mice. **a**, *sanroque* pedigree with ethylnitrosourea-treated founder (arrow) and ANA⁺ mice (green). **b**, **c**, Indirect immunofluorescence for antinuclear antibodies (HEp-2) (**b**) and dsDNA antibodies (*Crithidia luciliae*) (**c**) in *sanroque* sera. **d–g**, Histology of *sanroque*: IgG immunofluorescence (**d**) and light microscopy (haematoxylin/eosin stain) (**e**); liver (**f**); spleen size (left, *san/san*; right, *+/+*) (**g**). **h**, Lymphadenopathy. **i**, Thrombocytopenia (filled symbols, *san/san*; open symbols, *+/+*). **j**, Thrombocytopenia 2 days after passive transfer of *san/san* sera but not *+/+* sera (symbols as in **i**) into *+/+* mice. Each symbol represents one mouse. **k–n**, Plasmacytosis (syndecan-1, blue) in lymph node medullary cords of 22-week-old *sanroque* mice (**k**), and (methyl-pyronin-green, pink) in kidney (**l**), liver (**m**) and lung (**n**). **o**, Hypergammaglobulinaemia in 11-week-old mice (filled symbols, *san/san*; open symbols, *san/+*).

lymph nodes (Fig. 1g, h). Lymph node medullary cords are engorged with plasma cells (Fig. 1k) and plasmacytic infiltrates occur in kidneys, liver and lung (Fig. 1l–n). *sanroque* mice develop polyclonal hypergammaglobulinaemia of all T-dependent isotypes, most strikingly IgE (Fig. 1o; Supplementary Fig. 1c, d). T-cell and B-cell development, and B-cell maturation in the periphery, all proceed normally (Supplementary Figs 2–4), but large numbers of CD44^{high}CD62L^{low} CD4⁺ and CD8⁺ T cells accumulate with age (Fig. 2a, b; Supplementary Figs 3k and 5a–c). Both naive (CD44^{low}) and memory/activated (CD44^{high}) *sanroque* CD4⁺ T cells express higher levels of the T-cell co-stimulatory receptor ICOS than their wild-type counterparts, and this is already detectable before weaning (Fig. 2c, d) and in CD4⁺ thymocytes (not shown).

To identify the primary dysregulated cell type in *sanroque*, we constructed bone marrow chimaeras containing a 50:50 mix of *san/san* and wild-type haematopoietic cells. Recipients of *sanroque* bone marrow mixtures developed an expanded CD44^{high} T-cell population derived predominantly from the Ly5^b-marked cells bearing the *sanroque* mutation, establishing that the gene defect acts T-cell-autonomously (Fig. 2e, f; Supplementary Fig. 5e). Similarly, mutation-bearing naive and memory/activated CD4⁺ T cells

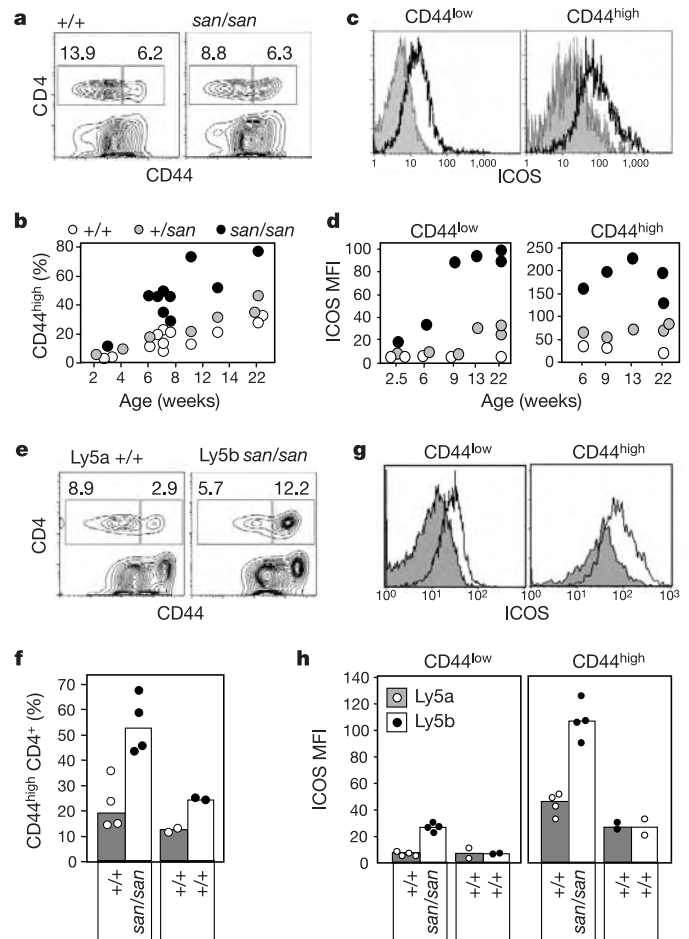


Figure 2 | Intrinsic T cell abnormalities. **a**, Flow cytometry of splenocytes from 6-week-old mice. **b**, Percentage of CD4⁺ spleen cells that are CD44^{high}. **c**, ICOS on naive (CD44^{low}) and activated (CD44^{high}) spleen CD4⁺ cells from *+/+* (shaded) and *san/san* (open) mice. **d**, ICOS mean fluorescence intensity (MFI) of CD4⁺CD44^{low} and CD4⁺CD44^{high} spleen cells. **e–h**, Irradiated B6.SJL-Ly5^a mice 12 weeks after reconstitution with a 50:50 mix of B6-Ly5^b *san/san* and B6.SJL-Ly5^a *+/+* bone marrow or a 50:50 mix of B6-Ly5^b *+/+* and B6.SJL-Ly5^a *+/+* bone marrow: CD44 on CD4⁺ blood leucocytes (**e**, **f**); ICOS on Ly5^a (*+/+*, shaded) and Ly5^b (*san/san*, open) CD4⁺CD44^{low} or CD4⁺CD44^{high} cells (**g**, **h**).

had higher levels of ICOS than wild-type *Ly5^a*-marked CD4⁺ T cells in the same mice (Fig. 2g, h). In contrast, B-cell hyperactivation (Supplementary Fig. 3i) and plasmacytosis are probably consequences of the T cell abnormalities, because a similarly elevated proportion of *san/san:Ly5^b* and *+/+ :Ly5^a* lymph-node B cells had high levels of CD69 activation antigen or the plasma cell marker syndecan-1 (not shown).

Stimulation of *sanroque* T cells with CD3 *in vitro* induced the activation markers CD69 and CD25 normally, but there was much greater induction of ICOS than on wild-type controls (Fig. 3a) and this was already evident after 6 h of activation (Supplementary Fig. 6a). Increased ICOS protein on *san/san* T cells was matched by a comparable increase in ICOS mRNA (Supplementary Fig. 6b). Although *sanroque* lymph-node T cells proliferate more than wild-type T cells in response to anti-CD3 (Supplementary Fig. 6c), when wild-type (*Ly5^a*) and *sanroque* (*Ly5^b*) T cells were co-cultured, *sanroque* T cells showed no intrinsic proliferative advantage (Supplementary Fig. 6d), and both followed a normal dose response to stimulation with anti-T-cell antigen receptor (anti-TCR) (Supplementary Fig. 6e) and made similar quantities of IL-2. By contrast, *san/san* T cells produced much more IL-5 and IFN- γ after stimulation (Fig. 3b).

Specific dysregulation of self-reactive T cells

To determine whether autoimmunity in *sanroque* mice simply reflects a general excess of T-cell helper activity directed indiscriminately to B cells specific for both self and foreign antigens, *sanroque* and wild-type mice were immunized with chicken gamma-globulin and *Bordetella pertussis*, which elicit antigen-specific Th2 and Th1 responses, respectively^{19,20}. The *sanroque* mutation did not alter the quantity or quality of the antibody response to these foreign antigens (Supplementary Fig. 6f).

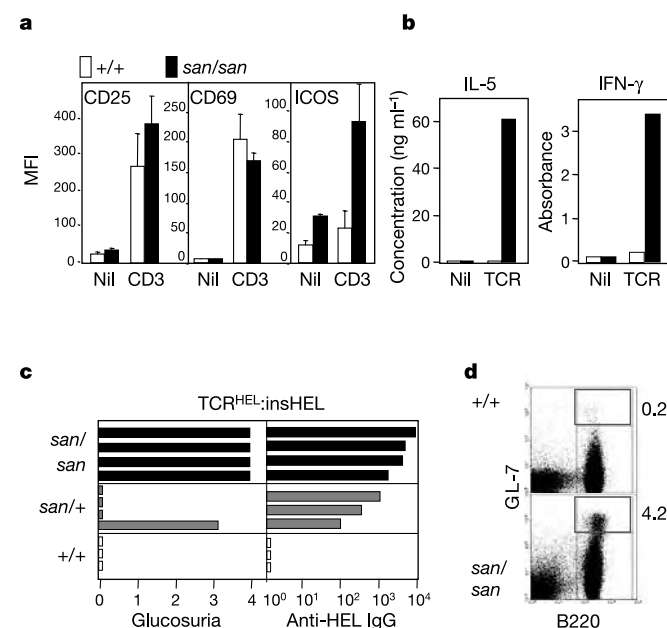


Figure 3 | Dysregulated T cell responses to TCR stimuli. **a**, Flow cytometry for CD25, CD69 and ICOS on purified CD4⁺ cells 24 h after incubation with plate-bound anti-CD3 (10 μ g ml⁻¹). Filled bars, *san/san*; open bars, *+/+*. MFI, mean fluorescence intensity. Error bars indicate standard deviation. **b**, IL-5 and IFN- γ in 6-day culture supernatants of lymph node cells stimulated with plate-bound anti-TCR (10 μ g ml⁻¹). Absorbance for IFN- γ was measured at 490 nm, with a reference wavelength of 405 nm. Filled bars, *san/san*; open bars, *+/+*. **c**, Diabetes and anti-islet autoantibody formation in 4–8-week-old H-2^{k/b} TCR^{HEL}:insHEL mice. **d**, Flow cytometry of pancreatic lymph nodes from 6-week-old TCR^{HEL}:insHEL transgenic mice for germinal centre (B220⁺ GL7⁺) cells.

To confirm a specific defect in the repression of self-reactive helper T cells, *sanroque* mice were crossed with TCR^{HEL}:insHEL double transgenic mice. In *+/+ :TCR^{HEL}:insHEL* mice, large numbers of CD4⁺ T cells reactive to hen-egg lysozyme (HEL) in the pancreatic islets are prevented from causing diabetes or anti-HEL antibodies because HEL—expressed as a self protein from the insulin gene promoter—triggers clonal deletion and anergy within HEL-reactive CD4⁺ T cells but not in B cells^{21,22}. By contrast, all *san/san:TCR^{HEL}:insHEL* mice were diabetic at weaning (Fig. 3c) and a proportion of *san/+ :TCR^{HEL}:insHEL* mice also developed diabetes after 8 weeks of age. *san/san:TCR^{HEL}:insHEL* mice produced high titres of anti-HEL IgG autoantibodies (Fig. 3c) and formed large populations of germinal-centre B cells in the draining pancreatic lymph nodes (Fig. 3d), indicating a profound failure to control T-cell help to self-reactive B cells in *sanroque* mice.

We next tested *sanroque* mice for defects in established mechanisms for controlling self-reactive T cells. Unlike *Aire*-deficient TCR^{HEL}:insHEL mice²³, no defect in thymic deletion of CD4⁺ cells was observed in *san/san:TCR^{HEL}:insHEL* animals (not shown). Unlike *Bim*-deficient mice²⁴, thymic deletion of self-reactive T cells bearing V β 11 TCRs in mice presenting endogenous *Mtv* super-antigen with I-E histocompatibility molecules²⁵ was normal in *sanroque* (Supplementary Fig. 7a). In contrast to *ctlb* deficiency, which renders mature T cells independent of CD28 co-stimulation and resistant to anergy induction^{26,27}, *sanroque* T cells showed normal CD28 dependence for efficient proliferation (Supplementary Fig. 7b). *sanroque* mice did not display the double-negative (CD90⁺CD4⁻CD8⁻) T-cell population characteristic of *lpr* and *gld* mice (Supplementary Fig. 7c), and FasL-dependent activation-induced T-cell death was intact in *san/san* CD4⁺ T cells (Supplementary Fig. 7d). CD40L expression on CD3-stimulated mature T cells, which is exaggerated in *NFATc1/c2* double-knockout mice²⁸, was normal in *sanroque* T cells (Supplementary Fig. 7e). Last, there was no deficiency in the total cell number or function of regulatory CD4⁺CD25⁺ T cells that are missing from *FoxP3*-mutant mice^{29,30} (Supplementary Fig. 7f–j). Taken together, these data indicate that the failure to repress autoimmune T cell help in *sanroque* mice does not correspond to defects in any of the known tolerance mechanisms.

Dysregulated follicular helper T cells

Unimmunized 8-week-old *sanroque* mice have large germinal centres in almost every follicle (Fig. 4a, b). Germinal-centre B cells increase with age, yielding a total population of about 2×10^7 cells in lymph nodes of 22-week-old mice (Supplementary Figs 2h, i and 3d, k). Strikingly, in *sanroque* mice, many splenic T cells are located within germinal centres and this is the predominant location of T cells in the spleens of older *san/san* mice (Fig. 4c, d; Supplementary Fig. 4). Germinal centres are 400% enriched in irradiated recipients reconstituted with *sanroque* bone marrow, compared with those reconstituted with wild-type bone marrow (Fig. 4e, f). Mixed chimaeras still showed an increase in *san/san* donor-derived (*Ly5^b*) germinal-centre B cells compared with both the number of control wild-type-derived (*Ly5^a*) germinal-centre B cells in the same mice (*U*-test, $P = 0.025$) and to chimaeras with *+/+ Ly5^b* donors (Fig. 4f), indicating that the excessive germinal-centre activity is not corrected by wild-type T or B cells, and that Roquin might also act cell-autonomously within germinal-centre B cells.

T-cell dysregulation was studied by comparing gene expression in purified *san/san* and *+/+* CD4⁺ T cells. Genes expressed more highly in *sanroque* CD4⁺ T cells included *Icos*, *Cxcr5*, *Cd200*, *Cd84*, *Ascl2*, *CXCL13*, *Pdcd1*, *Tiam1*, *neuropilin* and *Ccl5/RANTES* (Fig. 4g; Supplementary Fig. 9). Of 20 interleukin genes analysed, only that encoding IL-21 was expressed significantly more highly (more than fourfold). We confirmed the high levels of PDCD1, CXCR5, CD200, CCL5 and IL-21 by flow cytometry and real-time polymerase chain reaction (PCR) on sorted CD4⁺ cells (Fig. 4h, i). This constellation of overexpressed genes encode recently defined markers of the T_{FH} cell

for binding SH3 domains of interacting proteins, and two coiled-coil domains that could mediate homo-multimerization or hetero-multimerization. On the basis of the unique combination of RING and CCCH (C3H) domains, the gene symbols assigned in mouse and human are *Rc3h1* and *RC3H1*, respectively.

Western blotting for Roquin revealed the predicted 125-kDa Roquin protein in wild-type and *san/san* T lymphocytes, indicating that the mutation does not destabilize the protein (Fig. 6a). Confocal microscopy of 293T cells transfected with green fluorescent protein

(GFP)-tagged wild-type or mutant Roquin constructs showed that both proteins localized to distinct cytoplasmic dots (Fig. 6b) that co-stained for T-cell-induced antigen 1 (TIA-1) (Fig. 6c), a translational silencing factor and component of cytoplasmic stress granules, which are sites for regulating mRNA translation and decay³⁷.

Finally, we confirmed that the cell-autonomous dysregulation of *sanroque* T cells was corrected by wild-type Roquin. Full-length wild-type *roquin* cDNA was subcloned into a retroviral expression vector, upstream of an internal ribosomal entry site and enhanced green fluorescent protein (eGFP), and transduced into activated *sanroque* and wild-type T cells. ICOS overexpression in *sanroque* T cells was corrected in cells transduced with the wild-type *roquin* vector (Fig. 6d). Forced retroviral expression of Roquin^{M199→R} yielded an intermediate repression of ICOS on *san/san* T cells (Supplementary Fig. 12), which is consistent with a hypomorphic rather than null

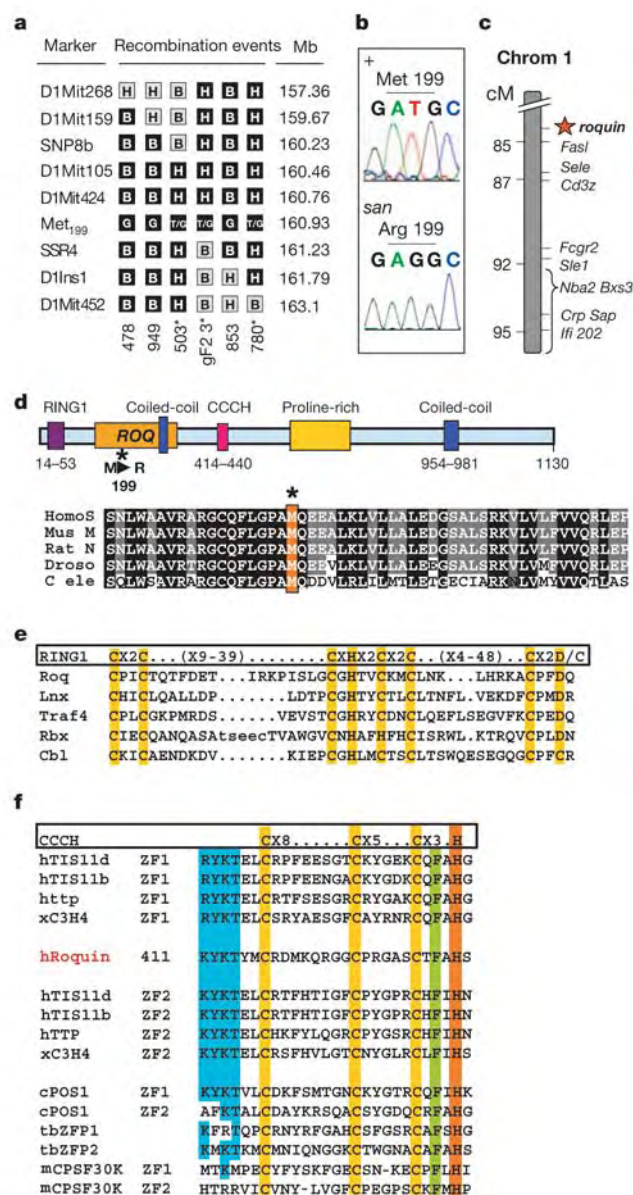


Figure 5 | Missense mutation in a conserved gene of previously unknown function. **a**, Meiotic mapping of *sanroque* mutation. Black squares represent B6 homozygosity in affected animals or B6/CBA heterozygosity in unaffected (asterisked) animals. **b**, T → G substitution in *roquin* mRNA base 596. **c**, Lupus genes in telomeric chromosome 1. **d**, Roquin primary structure, and conservation of ROQ domain and M199. **e**, Alignment of RING1 consensus (Pf00097) with Roquin RING finger and other RING-type E3 ubiquitin ligases. Gold highlights indicate zinc-coordinating residues in crystal structures of RBX1 and CBL. **f**, Alignment of Roquin CCCH zinc-finger with zinc-fingers (ZF) 1 and 2 from the TIS11 family proteins, human TTP, XC3H-4 (AAD24210), worm (*C. elegans*) POS-1, trypanosome tbZFP1 and tbZFP2, and mouse CPSF30K. Blue, conserved motifs (R/KYKT); gold and red, zinc binding; green, residue involved in base-stacking interactions with RNA.

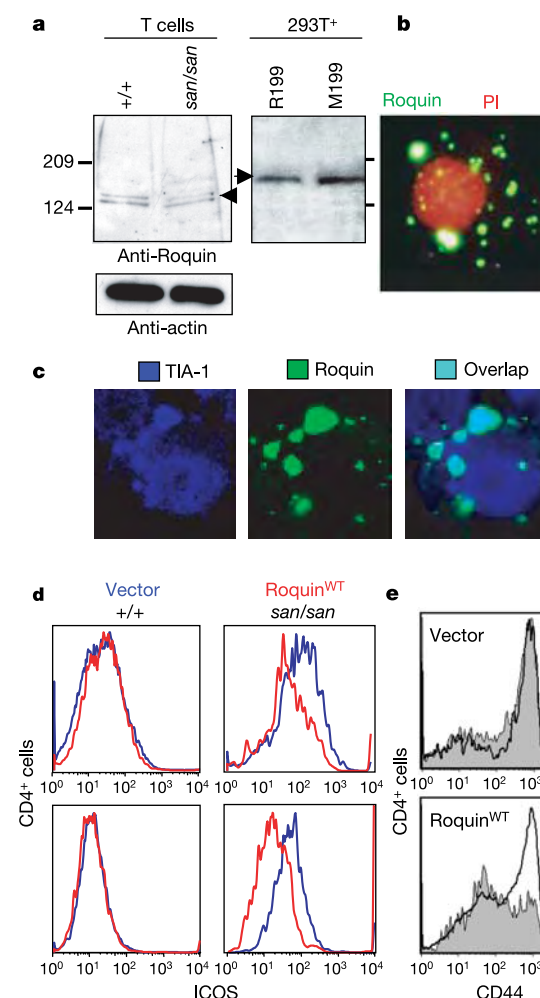


Figure 6 | Expression and function of Roquin protein. **a**, Western blots probed for Roquin (left, upper panel) or actin (left, lower panel) in thymocytes, or for Roquin (right) in 293T cells transfected with GFP-tagged wild-type (M199) or *sanroque* (R199) Roquin. **b**, **c**, Roquin-GFP localization to cytoplasmic granules in 293T cells: **b**, nucleus counterstained with propidium iodide (PI); **c**, granules counterstained for TIA-1. **d**, ICOS on GFP + *san/san* or wild-type T cells stimulated for 4 days (upper) or 6 days (lower) with plate-bound anti-CD3 (1 μ g ml⁻¹), anti-CD28 (10 μ g ml⁻¹) and IL-2 (100 U ml⁻¹), transduced with a bicistronic retroviral vector encoding wild-type Roquin and GFP (red) or GFP only (blue). **e**, CD44 on CD4⁺ T cells differentiating from *san/san* Ly9⁻ haematopoietic stem cells transduced (GFP +) with wild-type (WT) Roquin or control vector, compared to non-transduced (GFP⁻) CD4⁺ *san/san* cells in the same animals, 36 weeks after reconstitution of irradiated Ly9.1⁺ wild-type mice.

mutation. To test for gene complementation *in vivo*, wild-type Roquin-GFP or GFP-only retroviral vectors were transduced into *san/san* haematopoietic stem cells, and these were used to reconstitute the blood and lymphoid systems of irradiated wild-type recipients. CD44^{high} *san/san* T cells accumulated in the recipient animals as observed previously, and this was unaltered by transduction with the GFP-only control vector but was corrected specifically in GFP⁺ cells transduced with the *roquin* vector (Fig. 6e).

Discussion

By developing a systematic screen for autoimmune regulators in mice, we have identified a novel RING-type ubiquitin ligase family member, Roquin, that serves a critical role in preventing immune responses to self antigens. Prevention of autoimmunity by wild-type Roquin does not involve any known self-tolerance mechanisms but indicates a specific defect in selection of self-reactive B cells by T_{FH} cells. ICOS is overexpressed in *san/san* T cells and this is likely to be instrumental in the autoimmune phenotype, consistent with recent evidence that blockade of ICOS inhibits lupus in NZB/W F₁ mice³⁸. Quantitative differences in ICOS expression are implicated in immune abnormalities, because heterozygous ICOS-deficient mice have an intermediate depression of antibody responses³, whereas elevated ICOS expression occurs in human lupus³⁹ and in mice carrying a diabetes-susceptible *Idd5.1* haplotype of the ICOS gene⁴⁰. Similarly, excessive production of IL-21 has recently been associated with type 1 diabetes and with the inheritance of an autoimmune-susceptible *Idd3* haplotype of the *Il21* gene⁴¹.

Conversion of activated extrafollicular CD4⁺ T cells to T_{FH} cells requires bacterial adjuvants and CD40L/OX40L co-stimulation. They fail to differentiate under conditions favouring T-cell anergy^{42–44}. Roquin's critical role in repressing anti-self antibody responses might involve setting an inhibitory threshold that bars self-reactive T cells in the extrafollicular subset from assuming a follicular helper state. For example, T cells with low avidity for self antigen escape *Aire*-dependent thymic deletion in TCR^{HEL}:insHEL double transgenic animals but are normally anergic and very poor helpers for antibody responses²¹. Nevertheless, these cells clearly acquire potent T helper function when Roquin is defective (Fig. 3c, d).

Although the Roquin sequence clearly places it in the RING-type E3 ubiquitin ligase protein family, the substrate specificity of other members of this family, such as CBL or RBX1, is not determined by the E2 ubiquitin-binding RING domain but by unique domains in these proteins that recruit and position specific substrates by means of multi-protein complexes^{45,46}. In Roquin, the combination of a RING domain and an RNA-binding CCCH zinc-finger is distinctive and strongly indicates RNA-associated substrates. Temporal and spatial regulation of mRNA stability and translation by specific combinations of mRNA binding proteins is emerging as a key process in developmental patterning, synaptic plasticity, and inflammation^{33,37,47}. The mRNAs for essential T_{FH} molecules such as ICOS, IL-21 and SAP⁴⁸ all contain AUUUA sequences involved in mRNA degradation and translational repression. AUF1 and HuR bind to these sequences⁴⁹, and AUF1 stimulates the rapid turnover of AU-rich cytokine mRNAs through a poorly understood dependence on ubiquitination⁵⁰. These data, in conjunction with localization to TIA-1⁺ cytoplasmic RNA granules, makes Roquin a compelling candidate for regulating the ubiquitin-dependent control of AU-rich mRNAs by AUF1 and HuR.

Given Roquin's ubiquitous expression, its conservation in invertebrates, and preliminary evidence that a *Roquin*-null mutation in *Drosophila* is larval-lethal (R. Saint, personal communication), it is likely to serve wider developmental roles. It will therefore be important to test whether a null allele has additional consequences beyond the selective T-cell effects revealed by the *sanroque* missense mutation. Future work will need to explore the possibility that *roquin* allelic variants contribute to human autoimmune diseases. Understanding how *roquin* represses T_{FH} cells might illuminate strategies to

enhance T-cell responses to vaccines and to treat autoimmune diseases.

METHODS

Mice and procedures. Ethylnitrosourea treatment, construction of bone marrow chimaeras, and immunizations were as described previously²⁰. For colitis experiments, 4×10^5 sorted naive (CD4⁺CD25⁻) T cells with or without 4×10^5 sorted regulatory (CD4⁺CD25⁺) T cells were injected intravenously into RAG-1^{-/-} recipients. At week 5, 1-cm segments of proximal and distal colon were fixed in formalin and stained with haematoxylin and eosin. All mice were housed in specific pathogen-free conditions and all animal procedures were approved by the Australian National University Animal Ethics and Experimentation Committee.

Detection of autoantibodies. For ANAs, dilute plasma (dilution 1:40) was incubated on Hep-2 substrate. For anti-dsDNA antibodies, plasma (dilution 1:20) was incubated on *Crithidia luciliae* substrate. Bound antibody was detected with anti-mouse IgG conjugated with fluorescein isothiocyanate (FITC). Immune complex deposition was detected by staining frozen kidney sections with anti-mouse IgG-FITC.

Full blood counts. These were performed with an Abbott 3700 haematology analyser.

Serum electrophoresis. Serum proteins were separated by agarose-gel electrophoresis, stained with Coomassie blue and subjected to densitometry.

Tissue culture and antibody enzyme-linked immunosorbent assay. These were performed as described previously²⁰.

Flow cytometry, CFSE labelling and immunohistochemistry. These are described in Supplementary Methods.

RNA analysis. RNA was isolated from purified CD4⁺ T cells from two +/+ and *san/san* mice by using TRIzol. MOE 430A arrays (Affymetrix) were used and the samples were independently hybridized in duplicate in accordance with the manufacturer's instructions. Data were exported from GCOS and analysed in GeneSpring version 7.0. All measurements less than 0.01 were set to 0.01. The data were normalized such that the signal of each gene was divided by the median intensity of the chip and then each gene was divided by the median of its measurements in all samples. For real-time RT-PCR, total RNA was prepared from purified T cells by using TRIzol and reverse-transcribed with oligo(dT). cDNA expression was determined with the ABI Prism 7700 sequence detection system and TaqMan Syber-green reagents (PE Biosystems). Primer sequences are available on request. Fluorescence signals were measured over 40 PCR cycles and the cycle (*Ct*) at which signals crossed a threshold set within the logarithmic phase was recorded. The *Ct* for the target gene was subtracted from the *Ct* for β -actin (ΔCt). The relative amount of mRNA was calculated as $2^{\Delta Ct}$.

Mapping and sequencing. DNA from F₂ intercross mice was analysed by PCR amplification of SSLP or SNP markers. Primer sequences for new polymorphic markers and polymorphism data are shown in Supplementary Fig. 6a and will be deposited in the Mouse Genome Informatics website. Spleen and lymph-node cDNA were amplified by PCR and sequenced with primers complementary for all transcripts predicted by Ensembl to be in the interval.

Protein tagging, cell transfection and protein analysis. The coding sequence of *roquin* mRNA was PCR-amplified from B6 spleen cDNA, A-tailed with *Taq* DNA polymerase and cloned in-frame with the green fluorescence protein in the pCDNA3.1 CT-GFP TOPO TA fusion vector, to generate plasmid Roquin-GFP. The Met 199 $\rightarrow$ Arg substitution was introduced by site-directed mutagenesis. DNA coding for the wild-type and mutant Roquin was subcloned from Roquin-M¹⁹⁹-GFP and mutant Roquin-R¹⁹⁹-GFP plasmids and cloned between the *Xho*I and *Not*I sites of pMZS3F, thus fusing the C terminus of Roquin to the Flag epitope. Roquin-M¹⁹⁹-GFP and mutant Roquin-R¹⁹⁹-GFP were transiently transfected into 293T cells with the use of calcium phosphate. Cell lysates from transduced cells were prepared in TN buffer and protease inhibitors, and fractionated by SDS-polyacrylamide-gel electrophoresis. Membranes were probed with an affinity-purified rabbit polyclonal antibody against a C-terminal peptide of Roquin, followed by horseradish-peroxidase-conjugated goat anti-rabbit IgG and developed with enhanced chemiluminescence. Blots were reprobbed with anti-actin antibodies to control for protein loading. To study intracellular localization, transduced cells were fixed on coverslips (3% w/v paraformaldehyde in PBS) and permeabilized with 1% Triton X-100. Cells were stained with goat anti-TIA-1; after three washes they were incubated for 1 min with anti-goat-Alexa 493 with or without 0.75 mM propidium iodide counterstain and observed under fluorescence and confocal microscopes.

Retroviral transduction. cDNA encoding wild-type Roquin was subcloned into the polylinker of the MSCV-LTR retroviral vector, which encodes GFP 3' of an internal ribosomal entry site of a bicistronic message, and retrovirus-containing

supernatant collected from transfected Phoenix cells. Splenic CD4⁺ T cells, purified from C57BL/6 +/+ or *san/san* mice by magnetic cell sorting, were prestimulated for 24 h with 5 µg ml⁻¹ anti-CD3ε, 2.5 µg ml⁻¹ anti-CD28 and 100 U ml⁻¹ IL-2 and then spinoculated with supernatant containing 4 µg ml⁻¹ Polybrene. For the gene rescue experiments *in vivo*, E16 fetal liver cells of (B6 × CBA)F2IC-Ly9.1⁻ *san/san* genotype were expanded in the presence of IL-3 (10 ng ml⁻¹), IL-6 (10 ng ml⁻¹) and stem cell factor (25 ng ml⁻¹) for 24 h, before being retrovirally transduced by spinoculation. The cells were then cultured in fresh medium containing the same concentration of cytokines for a further 24 h before a second spinoculation and then washed. Recipient (B6 × CBA)-Ly9.1⁺ F₁ mice were lethally irradiated and given about 6 × 10⁶ cells intravenously into the lateral tail vein.

Received 2 December 2004; accepted 15 March 2005.

- Wakeland, E., Liu, K., Graham, R. & Behrens, T. Delineating the genetic basis of systemic lupus erythematosus. *Immunity* **15**, 397–408 (2001).
- Mills, D. M. & Cambier, J. C. B lymphocyte activation during cognate interactions with CD4⁺ T lymphocytes: molecular dynamics and immunologic consequences. *Semin. Immunol.* **15**, 325–329 (2003).
- Tafari, A. et al. ICOS is essential for effective T-helper-cell responses. *Nature* **409**, 105–109 (2001).
- McAdam, A. J. et al. ICOS is critical for CD40-mediated antibody class switching. *Nature* **409**, 102–105 (2001).
- Dong, C. et al. ICOS co-stimulatory receptor is essential for T-cell activation and function. *Nature* **409**, 97–101 (2001).
- Dong, C., Temann, U. A. & Flavell, R. A. Cutting edge: critical role of inducible costimulator in germinal center reactions. *J. Immunol.* **166**, 3659–3662 (2001).
- Mak, T. W. et al. Costimulation through the inducible costimulator ligand is essential for both T helper and B cell functions in T cell-dependent B cell responses. *Nature Immunol.* **4**, 765–772 (2003).
- Wong, S. C., Oh, E., Ng, C. H. & Lam, K. P. Impaired germinal center formation and recall T-cell-dependent immune responses in mice lacking the costimulatory ligand B7–H2. *Blood* **102**, 1381–1388 (2003).
- Parrish-Novak, J. et al. Interleukin 21 and its receptor are involved in NK cell expansion and regulation of lymphocyte function. *Nature* **408**, 57–63 (2000).
- Ozaki, K. et al. A critical role for IL-21 in regulating immunoglobulin production. *Science* **298**, 1630–1634 (2002).
- MacLennan, I. C. Germinal centers. *Annu. Rev. Immunol.* **12**, 117–139 (1994).
- Walker, L. S., Gulbranson-Judge, A., Flynn, S., Brocker, T. & Lane, P. J. Co-stimulation and selection for T-cell help for germinal centres: the role of CD28 and OX40. *Immunol. Today* **21**, 333–337 (2000).
- Hutloff, A. et al. ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. *Nature* **397**, 263–266 (1999).
- Lohning, M. et al. Expression of ICOS *in vivo* defines CD4⁺ effector T cells with high inflammatory potential and a strong bias for secretion of interleukin 10. *J. Exp. Med.* **197**, 181–193 (2003).
- Moser, B., Schaerli, P. & Loetscher, P. CXCR5⁺ T cells: follicular homing takes center stage in T-helper-cell responses. *Trends Immunol.* **23**, 250–254 (2002).
- Chtanova, T. et al. T follicular helper cells express a distinctive transcriptional profile, reflecting their role as non-Th1/Th2 effector cells that provide help for B cells. *J. Immunol.* **173**, 68–78 (2004).
- Kim, C. H. et al. Unique gene expression program of human germinal center T helper cells. *Blood* **104**, 1952–1960 (2004).
- Grimbacher, B. et al. Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. *Nature Immunol.* **4**, 261–268 (2003).
- Toellner, K. M. et al. T helper 1 (Th1) and Th2 characteristics start to develop during T cell priming and are associated with an immediate ability to induce immunoglobulin class switching. *J. Exp. Med.* **187**, 1193–1204 (1998).
- Jun, J. et al. Identifying the MAGUK protein Carma-1 as a central regulator of humoral immune responses and atopy by genome-wide mouse mutagenesis. *Immunity* **18**, 751–762 (2003).
- Akkraraju, S. et al. A range of CD4 T cell tolerance: partial inactivation to organ-specific antigen allows nondestructive thyroiditis or insulinitis. *Immunity* **7**, 255–271 (1997).
- Lesage, S. et al. Failure to censor forbidden clones of CD4 T cells in autoimmune diabetes. *J. Exp. Med.* **196**, 1175–1188 (2002).
- Liston, A., Lesage, S., Wilson, J., Peltonen, L. & Goodnow, C. Aire regulates negative selection of organ-specific T cells. *Nature Immunol.* **4**, 350–354 (2003).
- Bouillet, P. et al. BH3-only Bcl-2 family member Bim is required for apoptosis of autoreactive thymocytes. *Nature* **415**, 922–926 (2002).
- Scherer, M. T., Ignatowicz, L., Pullen, A., Kappler, J. & Marrack, P. The use of mammary tumor virus (Mtv)-negative and single-Mtv mice to evaluate the effects of endogenous viral superantigens on the T cell repertoire. *J. Exp. Med.* **182**, 1493–1504 (1995).
- Chiang, Y. et al. Cbl-b regulates the CD28 dependence of T-cell activation. *Nature* **403**, 216–220 (2000).
- Jeon, M. S. et al. Essential role of the E3 ubiquitin ligase Cbl-b in T cell anergy induction. *Immunity* **21**, 167–177 (2004).
- Peng, S. L., Gerth, A. J., Ranger, A. M. & Glimcher, L. NFATc1 and NFATc2 together control both T and B cell activation and differentiation. *Immunity* **14**, 13–20 (2001).
- Khattry, R., Cox, T., Yasayko, S. A. & Ramsdell, F. An essential role for Scurfin in CD4⁺ CD25⁺ T regulatory cells. *Nature Immunol.* **4**, 337–342 (2003).
- Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4⁺CD25⁺ regulatory T cells. *Nature Immunol.* **4**, 330–336 (2003).
- Ozaki, K. et al. Regulation of B cell differentiation and plasma cell generation by IL-21, a novel inducer of Blimp-1 and Bcl-6. *J. Immunol.* **173**, 5361–5371 (2004).
- Yoshinaga, S. K. et al. T-cell co-stimulation through B7RP-1 and ICOS. *Nature* **402**, 827–832 (1999).
- Blackshear, P. J. Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem. Soc. Trans.* **30**, 945–952 (2002).
- Barabino, S. M., Ohnacker, M. & Keller, W. Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs. *EMBO J.* **19**, 3778–3787 (2000).
- Hudson, B., Martinez-Yamout, M., Dyson, H. & Wright, P. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nature Struct. Mol. Biol.* **11**, 257–264 (2004).
- Morking, P. A. et al. TcZFP1: a CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides *in vitro*. *Biochem. Biophys. Res. Commun.* **319**, 169–177 (2004).
- Anderson, P. & Kedersha, N. Stressful initiations. *J. Cell Sci.* **115**, 3227–3234 (2002).
- Iwai, H. et al. Involvement of inducible costimulator-B7 homologous protein costimulatory pathway in murine lupus nephritis. *J. Immunol.* **171**, 2848–2854 (2003).
- Hutloff, A. et al. Involvement of inducible costimulator in the exaggerated memory B cell and plasma cell generation in systemic lupus erythematosus. *Arthritis Rheum.* **50**, 3211–3220 (2004).
- Greve, B. et al. The diabetes susceptibility locus Idd5.1 on mouse chromosome 1 regulates ICOS expression and modulates murine experimental autoimmune encephalomyelitis. *J. Immunol.* **173**, 157–163 (2004).
- King, C., Ilic, A., Koelsch, K. & Sarvetnick, N. Homeostatic expansion of T cells during immune insufficiency generates autoimmunity. *Cell* **117**, 265–277 (2004).
- Kearney, E. R., Pape, K. A., Loh, D. Y. & Jenkins, M. K. Visualization of peptide-specific T cell immunity and peripheral tolerance induction *in vivo*. *Immunity* **1**, 327–339 (1994).
- Fillatreau, S. & Gray, D. T cell accumulation in B cell follicles is regulated by dendritic cells and is independent of B cell activation. *J. Exp. Med.* **197**, 195–206 (2003).
- Brocker, T. et al. CD4 T cell traffic control: *in vivo* evidence that ligation of OX40 on CD4 T cells by OX40-ligand expressed on dendritic cells leads to the accumulation of CD4 T cells in B follicles. *Eur. J. Immunol.* **29**, 1610–1616 (1999).
- Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**, 533–539 (2000).
- Zheng, N. et al. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**, 703–709 (2002).
- Wickens, M. & Goldstrohm, A. A place to die, a place to sleep. *Science* **300**, 753–755 (2003).
- Crotty, S., Kersh, E. N., Cannons, J., Schwartzberg, P. L. & Ahmed, R. SAP is required for generating long-term humoral immunity. *Nature* **421**, 282–287 (2003).
- Okamoto, S. et al. Expression of the SH2D1A gene is regulated by a combination of transcriptional and post-transcriptional mechanisms. *Eur. J. Immunol.* **34**, 3176–3186 (2004).
- Laroia, G., Cuesta, R., Brewer, G. & Schneider, R. J. Control of mRNA decay by heat shock-ubiquitin-proteasome pathway. *Science* **284**, 499–502 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff of the Australian Phenomics Facility for care, breeding and phenotyping of mice; A. Prins for histology; D. Webb for help with cytokine ELISAs; E. Kucharska for help with immunizations; and A. Murtagh and D. Buckle for help with sequencing. The work was supported by grants from the Wellcome Trust (UK), the Juvenile Diabetes Research Foundation, the National Health and Medical Research Council, and the National Institutes of Health. C.G.V. was the recipient of a Wellcome Trust International Prize Travelling Fellowship.

Author Information The mouse Roquin sequence has been deposited in GenBank under accession number AY948287. The gene symbols assigned by the mouse and human genome nomenclature committees are *Rc3h1* and *RC3H1*, respectively. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.G.V. (carola.vinuesa@anu.edu.au) or C.C.G. (chris.goodnow@anu.edu.au).

Origin of the orbital architecture of the giant planets of the Solar System

K. Tsiganis¹, R. Gomes^{1,2}, A. Morbidelli¹ & H. F. Levison^{1,3}

Planetary formation theories^{1,2} suggest that the giant planets formed on circular and coplanar orbits. The eccentricities of Jupiter, Saturn and Uranus, however, reach values of 6 per cent, 9 per cent and 8 per cent, respectively. In addition, the inclinations of the orbital planes of Saturn, Uranus and Neptune take maximum values of ~ 2 degrees with respect to the mean orbital plane of Jupiter. Existing models for the excitation of the eccentricity of extrasolar giant planets^{3–5} have not been successfully applied to the Solar System. Here we show that a planetary system with initial quasi-circular, coplanar orbits would have evolved to the current orbital configuration, provided that Jupiter and Saturn crossed their 1:2 orbital resonance. We show that this resonance crossing could have occurred as the giant planets migrated owing to their interaction with a disk of planetesimals^{6,7}. Our model reproduces all the important characteristics of the giant planets' orbits, namely their final semimajor axes, eccentricities and mutual inclinations.

The planetary migration discussed above is a natural result of planet formation. After the giant planets were formed and the circumsolar gaseous nebula was dissipated, the Solar System was composed of the Sun, the planets and a debris disk of small

planetesimals. The planets then started to erode the disk, by either accreting or scattering away the planetesimals. The planets migrated because of the exchange of angular momentum with the disk particles during this process^{6,7}. Numerical simulations⁸ show that Jupiter was forced to move inward, while Saturn, Uranus and Neptune drifted outward. The orbital distribution of trans-neptunian objects is probably the result of such planetary migration⁷, and suggests that Neptune probably started migrating well inside 20 AU while the disk was extended up to 30–35 AU (refs 9–11).

During migration, the eccentricities and mutual inclinations of the planets are damped because of their gravitational interaction with the disk particles, in a process known as dynamical friction¹². However, the planets' orbital periods also change. If initially the planets' orbits were sufficiently close to each other, it is likely that they had to pass through low-order mean motion resonances (MMRs), which occur when the ratio between two orbital periods is equal to a ratio of small integers. These resonance crossings could have excited the orbital eccentricities of the resonance crossing planets. We focus our investigation on the 1:2 MMR between Jupiter and Saturn, as it is the strongest resonance.

In all our simulations, we started with a system where the initial

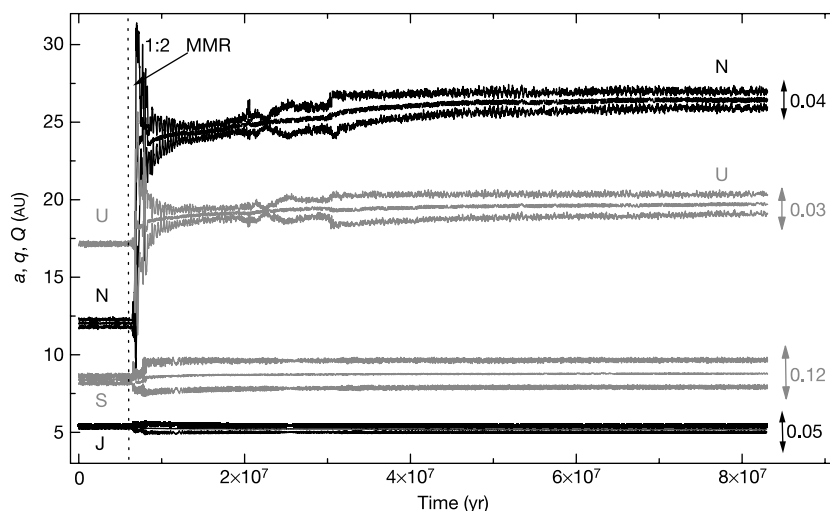


Figure 1 | Orbital evolution of the giant planets. These are taken from a N -body simulation with $35M_E$ 'hot' disk composed of 3,500 particles and truncated at 30 AU. Three curves are plotted for each planet: the semimajor axis (a) and the minimum (q) and maximum (Q) heliocentric distances. U, Uranus; N, Neptune; S, Saturn; J, Jupiter. The separation between the upper and lower curves for each planet is indicative of the eccentricity of the orbit. The maximum eccentricity of each orbit, computed over the last 2 Myr of

evolution, is noted on the plot. The vertical dotted line marks the epoch of 1:2 MMR crossing. After this point, curves belonging to different planets begin to cross, which means that the planets encounter each other. During this phase, the eccentricities of Uranus and Neptune can exceed 0.5. In this run, the two ice giants exchange orbits. This occurred in $\sim 50\%$ of our simulations.

¹Observatoire de la Côte d'Azur, CNRS, BP 4229, 06304 Nice Cedex 4, France. ²GEO/OV/UFRJ and ON/MCT, Ladeira do Pedro Antonio, 43-Centro 20.080-090, Rio de Janeiro, RJ, Brazil. ³Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Suite 400, Boulder, Colorado 80302, USA.

semimajor axis, a , of Jupiter was set to $a_J = 5.45$ AU and Saturn was placed a few tenths of an AU interior to the 1:2 MMR ($a_{1:2} \approx 8.65$ AU). The initial semimajor axes of the ice giants (Uranus and Neptune) were varied in the ranges 11–13 AU and 13.5–17 AU, while keeping their initial orbital separation larger than 2 AU. In all cases, the initial orbits of all the giant planets were nearly circular and coplanar (eccentricities, e , and mutual inclinations, i , $\sim 10^{-3}$). In addition to the giant planets, our simulations included a massive $((30\text{--}50)M_E$, where M_E is the mass of the Earth) particle disk, consisting of 1,000–5,000 equal-mass bodies, starting just beyond the orbits of the planets, ending between 30 and 35 AU, and with a surface density that falls linearly with heliocentric distance. It has been shown that, although this resolution is not enough to model all aspects of planetary migration¹¹, it adequately models the macroscopic evolution of the planetary orbits. Both dynamically ‘cold’ ($e \approx \sin i \approx 10^{-3}$) and dynamically ‘hot’ ($e \approx \sin i \approx 0.05$) disks were considered. We simulated the dynamical evolution of 43 different systems, using two different N -body codes, SyMBA¹³ and MERCURY¹⁴, with a time step of 0.25–0.5 years. In these experiments the self-gravity of the disk was ignored.

A typical example of the evolution undergone by our systems is shown in Fig. 1. At 6.6 Myr, after a period of slow migration on nearly circular orbits, Jupiter and Saturn cross the 1:2 MMR, at which point their eccentricities are quickly excited to values comparable to the ones currently observed. These ‘kicks’ in eccentricity are the result of the planets jumping over the 1:2 MMR without being trapped, and are qualitatively predicted by adiabatic theory (see Supplementary Information).

The sudden jump in the eccentricities of Jupiter and Saturn described above has a drastic effect on the planetary system as a whole, as shown in Fig. 1. The secular perturbations that Jupiter and Saturn exert on Uranus and Neptune force the eccentricities of the ice giants to increase by an amount that depends on the masses and semimajor axes of all planets¹⁵. As a result of the ‘compactness’ of the system, the planetary orbits become chaotic and intersect. When this occurs, a short phase of encounters follows the resonance crossing event. These encounters increase the inclinations of the planetary orbits by 1° – 7° . In addition, both ice giants are scattered outward and penetrate the disk. Thus, the flux of small bodies towards Saturn and Jupiter, and hence their rate of migration, increases abruptly. During this fast migration phase, the eccentricities and inclinations of the planets slowly decrease by dynamical friction and the planetary system is stabilized. The planets stop migrating when the disk is almost completely depleted. As shown in Fig. 1, not only their final semimajor axes, but also their final eccentricities, are close to the observed values.

The final orbits of the planets depend on the evolution of the system immediately after the resonance crossing event. Although there were many free parameters in our initial conditions, we found that the final configuration is most sensitive to the initial orbital separation between the ice giants ($\Delta a_{I_1, I_2}$) and, more importantly, to the one between Saturn and the inner ice giant ($\Delta a_{S, I_1}$). In our simulations, $\Delta a_{I_1, I_2}$ ranged from ~ 2 to ~ 6 AU, while $\Delta a_{S, I_1}$ ranged from ~ 2.5 to ~ 5 AU.

For $\Delta a_{S, I_1} < 3$ AU, the probability that Saturn scatters one of the ice giants to a Jupiter-crossing orbit increases. In such cases, the ice giant is ejected from the system. This happened in 14 (33%) of our runs. All other runs (67%) were successfully completed, that is, all four planets eventually reached stable orbits. Only two cases were found in which no encounters between the giant planets occurred. They both had $\Delta a_{S, I_1} \approx 5$ AU, which means that they were among the least compact systems that we simulated. In these runs, the semimajor axis of Uranus barely reached 16 AU, as in ref. 11. Repeated encounters between the ice giants were seen in all other successful runs. In 13 of them, only the ice giants encountered one another ($\Delta a_{S, I_1} \geq 3.5$ AU). For $\Delta a_{S, I_1} < 3.5$ AU, encounters between Saturn and an ice giant also occurred. Encounters with Saturn affect the dynamics of the

Jupiter–Saturn subsystem, allowing the gas giants to maintain their eccentricities against dynamical friction. This type of evolution was observed in 14 of our runs (33%). We note that, in this type of evolution, the duration of the fast migration phase is shorter than in the other cases.

Although we have not thoroughly explored the available parameter space, our experiments enable us to evaluate statistically the proposed excitation mechanism. We distinguish between two classes of runs: first, those in which there were no encounters between an ice giant and a gas giant (class A, 15 runs), and second, those in which Saturn suffered an encounter with one or both ice giants (class B, 14 runs). For each class, we computed the mean and standard deviation of the semimajor axis, proper eccentricity and proper inclination of each planet. Figure 2 shows the comparison between these quantities and the proper orbital elements of the real giant planets. Both classes of runs produce satisfactory results. Planetary orbits with very high eccentricities or inclinations are not produced. However, it is clear from this figure that class B runs ($\sim 50\%$ of our successful runs) give a much better match of the outer Solar System. In fact, the three orbital elements of all the real giant planets have values that lie within one standard deviation from the mean values of class B runs.

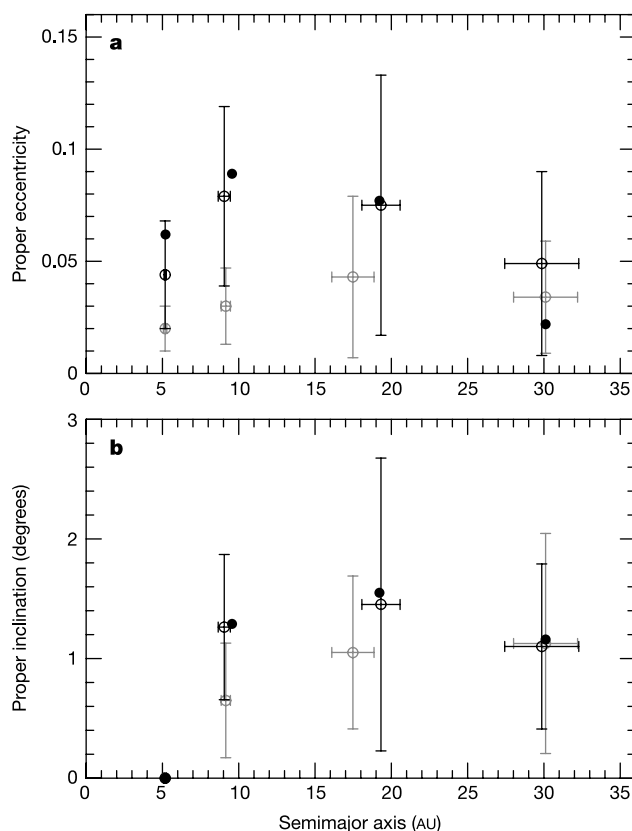


Figure 2 | Comparison of our synthetic final planetary systems with the outer Solar System. **a**, Proper eccentricity versus semimajor axis. **b**, Proper inclination versus semimajor axis. Proper eccentricities and inclinations are defined as the maximum values acquired over a 2-Myr timespan and were computed from numerical integrations. The inclinations are measured relative to Jupiter's orbital plane. These values for the real planets are presented with filled black circles. The open grey circles mark the mean of the proper values for the runs of class A (no encounters for Saturn), while the open black circles mark the same quantities for the runs of class B (see text for the definition of these classes). The error bars represent one standard deviation. The largest values of the proper eccentricity and inclination of our synthetic planets were $e = 0.11$ for Jupiter, $e = 0.17$ and $i = 2.5^\circ$ for Saturn, $e = 0.23$ and $i = 4.5^\circ$ for Uranus, and $e = 0.17$ and $i = 4.0^\circ$ for Neptune.

The final semimajor axes of the planets are an important diagnostic of migration models. The simulations of compact systems in ref. 11 always produced final configurations in which Neptune was at ~ 30 AU, but Uranus was too close to the Sun. Our model nicely solves this nagging problem. As shown in Fig. 2, class B runs give $a_U = 19.3 \pm 1.3$ AU and $a_N = 29.9 \pm 2.4$ AU, the observed values being $a_U = 19.2$ AU and $a_N = 30.1$ AU. (Here a_U and a_N are the semimajor axes of Uranus and Neptune, respectively.) The final orbital separation of Jupiter and Saturn depends on the amount of mass that they process during the evolution of the system — that is, on the initial mass of the disk. Although larger disk masses favour the stability of the four-planet system, we found that, for disk masses larger than $\sim (35\text{--}40)M_E$, the final orbital separation of Jupiter and Saturn tends to be larger than is actually observed. For disks of $50M_E$, Saturn was found to cross the 2:5 MMR with Jupiter. In addition, the final eccentricities of the two planets were too small, because they had experienced too much dynamical friction. Indeed, the fact that we reproduce both the semimajor axes and the eccentricities/inclinations in the same integrations is a strong point of our model.

The initial dynamical state of the disk also affects the final state of the planetary system. 'Hot' disks tend to produce systems where the eccentricities for Jupiter and Saturn are larger than in 'cold' disks. The actual disk may indeed have been as excited as we assumed in our 'hot' runs, because of the presence of a large number of Pluto-sized objects¹⁶.

Other compact planetary configurations could lead to the crossing of different MMRs. For reasons of completeness, we studied the crossing of the 2:3 and 1:2 MMRs between (1) Saturn and the inner ice giant, and (2) the two ice giants, by placing Saturn exterior to the 1:2 MMR with Jupiter, and varying the initial positions of Uranus and Neptune. We found that, although some of these resonance crossings may destabilize the orbits of the ice giants, none can excite the orbit of Jupiter.

The survivability of the regular satellites during the planetary encounters is a potential issue with our model. Thus, during eight migration simulations we recorded all encounters deeper than one Hill radius (approximately the distance within which the gravity of the planet dominates over the gravity of the Sun). We then integrated the evolution of the regular satellites of Saturn and the ice giants during a re-enactment of these encounters. We assumed that both ice giants had Uranus's satellite system. We found that in half of the simulations, all of the satellite systems survived the entire suite of encounters (that is, $\sin i, e < 0.05$). Thus, we conclude that the survivability of the satellites is not a problem for the model. However, we note that the irregular satellites would not survive the encounters. Thus, if this model is correct they must have been captured either during or after the 1:2 MMR crossing.

We noticed in our simulations that several particles were trapped on long-lived orbits characteristic of Neptune's Trojan asteroids (two per run, on average, with a lifetime larger than 80 Myr). Their eccentricities reached values < 0.1 . These particles were eventually removed from the Trojan region, but this is probably an artefact of the graininess of Neptune's migration⁸ (although this graininess could also have been responsible for their capture). Jupiter's Trojans are a more subtle issue, described in ref. 17, which also turns out to be a strength of our model.

Thus we conclude that the eccentricities of Jupiter and Saturn are

probably the result of the fact that these planets crossed the 1:2 MMR. Other mechanisms^{3–5} that have been proposed for the eccentricity excitation of extrasolar planets have neither been applied to our Solar System nor confronted with the large body of constraints that its current structure provides. Our model statistically reproduces all aspects of the orbits of the giant planets. It is consistent with the existence of regular satellites, with the observed distributions of Jupiter's Trojans¹⁷, perhaps with the existence of Neptune's Trojans, and does not contradict the distribution of main-belt asteroids¹⁸.

Received 6 December 2004; accepted 11 March 2005.

- Pollack, J. B. *et al.* Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **164**, 62–85 (1996).
- Lubow, S. H., Seibert, M. & Artymowicz, P. Disk accretion onto high-mass planets. *Astrophys. J.* **526**, 1001–1012 (1999).
- Goldreich, P. & Sari, R. Eccentricity evolution for planets in gaseous disks. *Astrophys. J.* **585**, 1024–1037 (2003).
- Papaloizou, J. C. B., Nelson, R. P. & Masset, F. Orbital eccentricity growth through disk-companion tidal interaction. *Astron. Astrophys.* **366**, 263–275 (2001).
- Zakamska, N. L. & Tremaine, S. Excitation and propagation of eccentricity disturbances in planetary systems. *Astron. J.* **128**, 869–877 (2004).
- Fernandez, J. A. & Ip, W.-H. Some dynamical aspects of the accretion of Uranus and Neptune—The exchange of orbital angular momentum with planetesimals. *Icarus* **58**, 109–120 (1984).
- Malhotra, R. The origin of Pluto's orbit: implications for the Solar System beyond Neptune. *Astron. J.* **110**, 420–432 (1995).
- Hahn, J. M. & Malhotra, R. Orbital evolution of planets embedded in a planetesimal disk. *Astron. J.* **117**, 3041–3053 (1999).
- Gomes, R. The origin of the Kuiper Belt high-inclination population. *Icarus* **161**, 404–418 (2003).
- Levison, H. F. & Morbidelli, A. The formation of the Kuiper belt by the outward transport of bodies during Neptune's migration. *Nature* **426**, 419–421 (2003).
- Gomes, R. S., Morbidelli, A. & Levison, H. F. Planetary migration in a planetesimal disk: Why did Neptune stop at 30 AU? *Icarus* **170**, 492–507 (2004).
- Kokubo, E. & Ida, S. Orbital evolution of protoplanets embedded in a swarm of planetesimals. *Icarus* **114**, 247–257 (1995).
- Duncan, M. J., Levison, H. F. & Lee, M. H. A multiple time step symplectic algorithm for integrating close encounters. *Astron. J.* **116**, 2067–2077 (1998).
- Chambers, J. E. A hybrid symplectic integrator that permits close encounters between massive bodies. *Mon. Not. R. Astron. Soc.* **304**, 793–799 (1999).
- Murray, C. & Dermott, S. F. *Solar System Dynamics* (Cambridge Univ. Press, Cambridge, UK, 1999).
- Stern, S. A. On the number of planets in the outer solar system—Evidence of a substantial population of 1000-km bodies. *Icarus* **90**, 271–281 (1991).
- Morbidelli, A., Levison, H. F., Tsiganis, K. & Gomes, R. Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* doi:10.1038/nature03540 (this issue).
- Gomes, R., Tsiganis, K., Morbidelli, A. & Levison, H. F. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* doi:10.1038/nature03676 (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements R.G. is grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico for financial support of his sabbatical year in the OCA observatory in Nice. The work of K.T. was supported by an EC Marie Curie Individual Fellowship. A.M. and H.F.L. thank the CNRS and the NSF for funding the collaboration between the OCA and the SwRI groups. H.F.L. is grateful to NASA's Origins and PG&G programmes.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M. (morby@obs-nice.fr).

LETTERS

Chaotic capture of Jupiter's Trojan asteroids in the early Solar System

A. Morbidelli¹, H. F. Levison^{1,2}, K. Tsiganis¹ & R. Gomes^{1,3}

Jupiter's Trojans are asteroids that follow essentially the same orbit as Jupiter, but lead or trail the planet by an angular distance of ~ 60 degrees (co-orbital motion). They are hypothesized to be planetesimals that formed near Jupiter and were captured onto their current orbits while Jupiter was growing^{1,2}, possibly with the help of gas drag^{3–6} and/or collisions⁷. This idea, however, cannot explain some basic properties of the Trojan population, in particular its broad orbital inclination distribution, which ranges up to ~ 40 degrees (ref. 8). Here we show that the Trojans could have formed in more distant regions and been subsequently captured into co-orbital motion with Jupiter during the time when the giant planets migrated by removing neighbouring planetesimals^{9–12}. The capture was possible during a short period of time, just after Jupiter and Saturn crossed their mutual 1:2 resonance, when the dynamics of the Trojan region were completely chaotic. Our simulations of this process satisfactorily reproduce the orbital distribution of the Trojans and their total mass.

Recent numerical experiments^{13,14} have shown that the orbits of the giant planets are best reproduced if Saturn and Jupiter crossed their mutual 1:2 mean motion resonance (MMR) during their migration. This occurs when the ratio of their orbital periods, P_S/P_J , equals 2. The current ratio of P_S/P_J is slightly less than 2.5. However, there is a serious argument in the literature against the idea that Jupiter and Saturn crossed the 1:2 MMR. If the crossing had happened, any pre-existing jovian Trojans would have become violently unstable, and Jupiter's co-orbital region would have emptied^{15,16}. Indeed, we performed a simulation similar to that in ref. 15, but with 1.3 million particles in the Trojan region—none survived the 1:2 MMR crossing.

However, the dynamical evolution of a gravitating system of objects is time reversible. Thus, if the local objects can escape the Trojan region when the latter becomes unstable, other bodies can enter the same region and be temporarily trapped. Consequently, a transient Trojan population can be created if there is an external source of objects. In this case, the source is constituted by the very bodies that are forcing the planets to migrate^{9–12}, and is of considerable magnitude given how much the planets must move. When Jupiter and Saturn get far enough from the 1:2 MMR that the co-orbital region becomes stable, the population that happens to be there at that time remains trapped. It becomes the population of permanent jovian Trojans still observable today.

To investigate the above idea, we first performed a numerical simulation that involved integrating the orbits of a series of massless planetesimals initially on Saturn-crossing orbits under the gravitational influence of the Sun, Jupiter and Saturn. In this simulation, the planets were on non-migrating orbits close to the 1:2 MMR, so that the Trojan region was fully unstable. We found that $\sim 1\%$ of the planetesimals initially on Saturn-crossing orbits spent more than

100 yr as jovian Trojans, which we define as objects having orbital periods relative to Jupiter's of between 0.97 and 1.03, absolute values of angular distance from the planet of between 40° and 90° , and eccentricities of less than 0.15. These values were derived from the current orbital distribution of the Trojans. The particles temporarily trapped in the Trojan region covered the whole region of co-orbital motion. More importantly, their orbital inclination covered all values up to $\sim 40^\circ$, as a result of previous close encounters with the planets. Therefore, our idea became appealing because it could potentially explain the puzzling broad inclination distribution of the jovian Trojans. Motivated by this possibility, we proceeded with a far more comprehensive, and time consuming, set of simulations of this idea.

The first step in this expanded study was to determine exactly when the Trojans become unstable during the resonant crossing. For this purpose, we started by adopting the migration rates from one of the simulations reported in ref. 13. In particular, we chose a simulation where the planets migrated relatively slowly. From that simulation we measured the ratio P_S/P_J at 40 timesteps (Fig. 1a). Then, we performed 40 orbital integrations of massless test particles under the influence of the Sun, Jupiter and Saturn. The planets were placed on non-migrating orbits, with the same values of P_S/P_J as measured in Fig. 1a. The initial distribution of test particles was chosen to mimic the current distribution of Trojans relative to Jupiter. Each simulation covered 2×10^5 yr, and the fraction of the initial test particle population that remained in the Trojan region is reported in Fig. 1b, where each simulation is represented by a single point. We note two planetary configurations that are critical for the survivability of the Trojans. One occurs when $P_S/P_J \approx 2.05$ (time $t = 4.5 \times 10^5$ yr in the reference simulation), at which point all resident Trojans escape. This indicates that the entire co-orbital region is particularly unstable at this time. This instability is due to a secondary 3:1 resonance¹⁷ between $(1/P_J - 2/P_S)$ and the oscillation frequency of the Trojans around the Lagrange point. The other critical configuration occurs when $P_S/P_J \approx 2.08$ ($t = 10^6$ yr), which corresponds to a secondary 2:1 resonance between the same two frequencies, and depletes 70% of the Trojans.

In our scheme, the capture of jovian Trojans had to have occurred during these two critical planetary configurations. Thus, we designed a pair of simulations intended to study the capture process. In the first of these simulations (referred to as the slow simulation hereafter), we adopted the same migration rate as in the last paragraph. Jupiter and Saturn were forced to migrate by including a suitably chosen drag term in the planets' equations of motion, as prescribed in ref. 10, so that they reproduced the evolution of P_S/P_J shown in Fig. 1a. From 3.5×10^5 yr (just before the first critical configuration is reached) to 1.2×10^6 yr (just after the second critical configuration has passed), we supplied a steady flux of 5,466,000 planetesimals through the Jupiter–Saturn system (see Methods). This simulation

¹Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France. ²Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Suite 400, Boulder, Colorado 80302, USA. ³GEA/OV/UFRJ and ON/MCT, Ladeira do Pedro Antonio, 43-Centro 20.080-090, Rio de Janeiro, RJ, Brazil.

covered 10 Myr, at which point the orbits of the planets were sufficiently close to their observed ones. The second simulation was identical to the first, but with a migration rate that was three times larger and an integration time three times shorter. We will refer hereafter to this as the fast simulation. Comparably fast migration rates have been observed in many of the runs in ref. 13.

At the end of the slow simulation, 2.4×10^{-6} of the planetesimals were found to be on orbits trapped in the Trojan region. The capture efficiency rises to 1.8×10^{-5} in the fast simulation. Of these trapped Trojans, $\sim 50\%$ (the same ratio in both simulations) have libration amplitudes (the semi-amplitude of the oscillation of the angular distance from Jupiter) smaller than 30° , like 87% of the known Trojans. The vast majority of the captured Trojans with larger amplitudes of libration would not survive up to current times, because their dynamics are unstable on long timescales¹⁸. Thus, we restrict the analysis of our fictitious Trojans to those objects with libration amplitudes less than 30° .

In terms of total mass (see Methods for a description of the mass estimates), our trapped Trojan population is quite consistent with the real population, when scaled to the mass required to move the planets the required distance. Our slow simulation predicts a total Trojan mass of $\sim 4 \times 10^{-6} M_E$ (where M_E is the Earth's mass), where the fast simulation predicts a mass of $\sim 3 \times 10^{-5} M_E$. Using the most up-to-date observations, we estimate the mass of the Trojan population with $D < 30^\circ$ to be $1.1 \times 10^{-5} M_E$. So, the actual mass of the Trojans appears to be in the range predicted by our two simulations.

The reason why the mass trapped in the Trojan region increases so sharply with the planetary migration rate is twofold. First, a faster migration rate corresponds to a proportionally higher mass flux. Thus, the transient population that resides in the co-orbital region

when this region is chaotic is proportionally enhanced. This explains a factor of about three between the results of the two simulations. Second, faster migration results in a sharper transition from instability to stability in the co-orbital region, which increases the fraction of the transient population that becomes permanently trapped. This probably explains the remaining factor of $\sim 7/3$ between the results of the two simulations.

Figure 2 shows the distribution of the captured Trojans in the space of the three fundamental orbital parameters that characterize co-orbital dynamics: the proper eccentricity e , inclination i and libration amplitude D . ('Proper' refers to parameters that are suitably averaged over short periods of time and is usually introduced to characterize oscillating orbits¹⁹.) Their computation is explained in Methods. Because the distribution of the captured objects is similar in both simulations, we have included both data sets in Fig. 2 in order to improve statistics. The distribution of the known Trojans is also plotted, for visual comparison. There is an excellent qualitative agreement between the observed and simulated distributions. The captured Trojans cover the same range of values of the orbital parameters as the observed ones. There is no macroscopic region of orbital parameter space that is both occupied by the real Trojans and is left empty by the simulated ones. In particular, simulated Trojans are found even on orbits with $D < 5^\circ$. These orbits are the hardest to populate, in any capture model⁸. We stress that the inclinations of the trapped Trojans range from 0° to 40° , like those of the observed population.

We note that our results may provide an explanation for why jovian Trojans look so similar to cometary nuclei and to some (the bluest) Centaurs and Kuiper belt objects at visible wavelengths^{20,21}. In fact, it has been argued that both the Kuiper belt^{22,23} and the scattered disk²⁴—which is the current source of Centaurs and Jupiter-family comets, and probably also the progenitor of the Oort cloud—

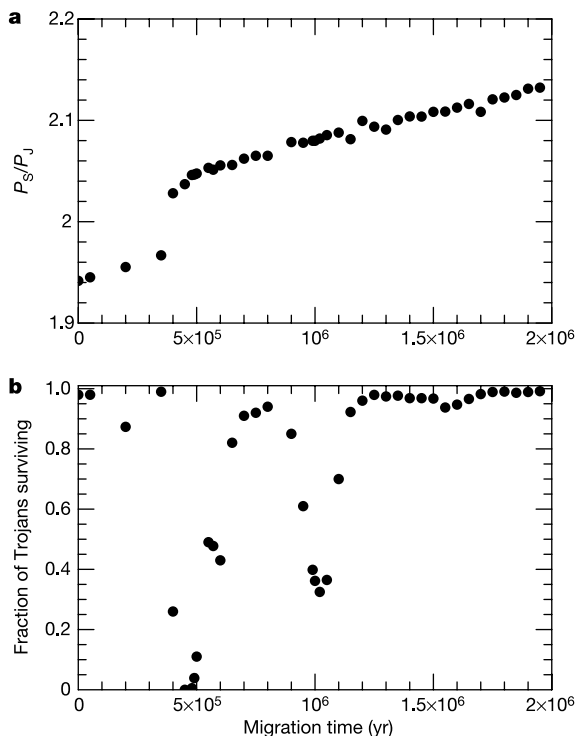


Figure 1 | The stability of Trojans during planetary migration. **a**, The temporal evolution of the ratio of orbital periods of Saturn and Jupiter (P_S and P_J , respectively) from the migration simulation that we chose from ref. 13. The magnitude of the jump in P_S/P_J when the planets cross $P_S/P_J = 2$ reflects the width of the 1:2 MMR. The planets are not captured into resonance but jump over it. **b**, The fraction of the Trojan population that survives for 2×10^5 yr in the co-orbital region, as a function of P_S/P_J (and hence of migration time).

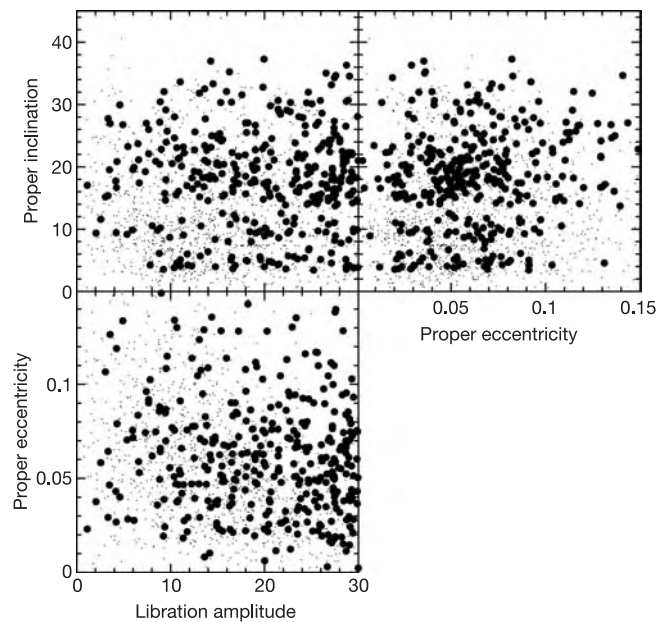


Figure 2 | Comparison of the orbital distribution of Trojans between model and observations. The simulation results are shown as filled circles and the observations as dots in the space of the three orbital parameters for co-orbital motion. The distribution of the simulated Trojans is somewhat skewed towards large libration amplitudes, relative to the observed population. However, this is not a serious problem because a fraction of the planetesimals with the largest amplitudes would leave the Trojan region during the subsequent 4 Gyr of evolution¹⁸, leading to a better match. The similarity between the two inclination distributions is strong support for our model. Libration amplitude and proper inclination are measured in degrees.

originated in the planetesimal disk that drove planetary migration. Our model places the origin of jovian Trojans in the same parent population.

Our results may also prove an explanation for the fact that Trojans are apparently deficient in water and organics²⁵. Before being captured in the Trojan region, planetesimals typically evolved through a large eccentricity phase that brought them relatively close to the Sun. Indeed, all the particles that spent more than 100 yr in the Trojan region in our first simulation reached a perihelion distance q of less than ~ 3 AU. Of them, 72% spent more than 10,000 yr on orbits with $q < 3$ AU, and 68% even reached $q < 2$ AU. Because it takes roughly 10,000 yr for an active Jupiter-family comet to become dormant²⁶, it is possible that the surfaces of the Trojans could have been devolatilized during their high eccentricity phase.

Our simulation shows that objects captured into Jupiter's co-orbital regions immediately after Jupiter and Saturn crossed the 1:2 MMR have an orbital distribution remarkably similar to that of the observed Trojans. In addition, it shows that the capture efficiency can explain the total number of objects observed. As our model is the only one available that can explain these features, we believe that the Trojans represent observational evidence for this resonance crossing, which has also been shown elsewhere¹³ to produce the correct planetary orbits. Thus, this work, together with refs 13 and 14, provides a self-consistent view of the formation and primordial evolution of the Solar System.

METHODS

Simulation of Trojan capture. We start with our 'slow' simulation, where Jupiter and Saturn are forced to migrate as in Fig. 1a. The flux of planetesimals is modelled by setting test particles on Saturn crossing orbits with orbital periods larger than P_S and a distribution of eccentricities and inclinations that mimic that in our reference simulation from ref. 13 when the 1:2 MMR is crossed. Every time that a test particle is dynamically eliminated, it is reintroduced on its original trans-saturnian orbit rescaled to the current position of Saturn. In this way, the number of particles in the simulation at any time is constant (1,163,000) and their orbital distribution remains in steady state. In total, 5,466,000 particles are eliminated and reintroduced during the considered time-span. At $t = 1.2 \times 10^6$ yr, when the co-orbital region becomes stable again, 98 particles are found on Trojan-like orbits. These particles are each cloned 19 times. The integration is then continued with the planets migrating, for 10 Myr, until the planets come reasonably close to their current semimajor axes. A drag force is also added to the planets' equations of motion, in order to slowly damp their eccentricities to their current values. At the end of the simulation, 266 particles are in the Trojan region. The final trapping efficiency is $266/20/5,466,000 \approx 2.4 \times 10^{-6}$.

In the 'fast' simulation, the migration rate of the planets is increased by a factor of three. A total of 2,773,000 particles are eliminated and reintroduced. 174 particles are found to be on Trojan-like orbits at the end of the second 'critical planetary configuration'. Of these particles, cloned 9 times each, 486 survive in the Trojan region at the end of planetary migration. The final trapping efficiency is $486/10/2,773,000 \approx 1.8 \times 10^{-5}$.

The above simulations did not take into account Uranus and Neptune. These planets could affect the capture of Trojans in two ways. Immediately after the 1:2 MMR crossing they provided kicks to Saturn during close encounters. Thus, we modify the first stage of the above simulations by including stochastic kicks to Saturn every 150,000 years with a magnitude of 0.53 km s^{-1} (based on ref. 13). Then during the post-capture, 10 Myr migration, Uranus and Neptune could destabilize the Trojans by generating additional resonances. Thus, we perform again the second stage simulation, but including Uranus and Neptune. These planets are forced to migrate from 16.5 and 20 AU to their current positions, while their eccentricities are damped from 0.1. We find that the inclusion of the ice giants does not affect Trojan capture.

Estimates of Trojan mass. According to ref. 13, $\sim 3.4 M_E$ of planetesimals are cycled through the system as the planets migrate through the unstable Trojan configurations. Of the trapped planetesimals, $\sim 50\%$ are in the region $D < 30^\circ$. The mass of captured Trojans is the product of $3.4 M_E$, 0.5, and the capture efficiency of the corresponding simulation.

According to ref. 27, the current mass of the Trojans is ~ 3 – 25 times larger than the value we find. However, ref. 27 probably overestimated the real value because it assumed: (1) an outdated density of 2 g cm^{-3} , whereas it is now

believed to be $\rho = 1.3 \text{ g cm}^{-3}$ (refs 28, 29); (2) an outdated mean albedo $p_v = 0.04$, whereas later observations²⁰ showed that it is probably $p_v = 0.056$, and (3) an absolute magnitude (H) distribution that predicts 2.5 times more objects with $H < 11$ than observed.

Correcting for (1) and (2), while keeping the H -distribution shown in Fig. 9 of ref. 2 reduces the Trojan mass estimate to $2.5 \times 10^{-5} M_E$. Correcting for (3) requires a more involved procedure. Reference 27 constrains the slope of the H -distribution for $H > 10.5$, but their estimate of the total number is problematic because of the paucity of bright objects observed in their narrow field deep survey. To overcome this problem, we use the most recent catalogue of Trojan bodies (<http://Hamilton.unipi.it/cgi-bin/astdys/astibo>) which, according to SDSS findings (Gy. M. Svabo and Z. Ivezić, personal communication) is complete up to $H = 11.5$. Beyond this threshold we extrapolate the catalogue's distribution using the slope given in ref. 27. This reduces the total mass of the Trojans to $1.3 \times 10^{-5} M_E$, 87% of which is in the considered $D < 30^\circ$ region.

Computation of Trojan proper elements. We integrate each Trojan orbit for 10^5 yr under the gravitational influence of only the Sun and Jupiter. No planetary migration is imposed. The numerical output is digitally filtered³⁰ in order to eliminate the short periodic oscillations of the orbital elements. The libration amplitude D is computed as $(\delta\lambda_{\max} - \delta\lambda_{\min})/2$, where $\delta\lambda$ is the difference between the mean longitude of the Trojan and of Jupiter, and the suffices min and max denote, respectively, its minimal and maximal value over a libration cycle. The proper eccentricity is computed as $(k_{\max} - k_{\min})/2$, where $k = e \sin \varpi$, ϖ is the Trojan's perihelion longitude and $k_{\max/\min}$ are computed over a secular oscillation of the Trojan's orbit. The proper inclination is computed in a similar way. This procedure is consistent with that used in ref. 19 for the real Trojans, which allows a direct comparison in Fig. 2.

Received 6 December 2004; accepted 11 March 2005.

- Marzari, F. & Scholl, H. Capture of Trojans by a growing proto-Jupiter. *Icarus* **131**, 41–51 (1998).
- Fleming, H. J. & Hamilton, D. P. On the origin of the Trojan asteroids: Effects of Jupiter's mass accretion and radial migration. *Icarus* **148**, 479–493 (2000).
- Yoder, C. F. Notes on the origin of the Trojan asteroids. *Icarus* **40**, 341–344 (1979).
- Peale, S. J. The effect of the nebula on the Trojan precursors. *Icarus* **106**, 308–322 (1993).
- Kary, D. M. & Lissauer, J. J. Nebular gas drag and planetary accretion. II. Planet on an eccentric orbit. *Icarus* **117**, 1–24 (1995).
- Kortenkamp, S. J. & Hamilton, D. P. Capture of Trojan asteroids in the early Solar Nebula. *Bull. Am. Astron. Soc.* **33**, 1086 (2001).
- Shoemaker, E. M., Shoemaker, C. S. & Wolfe, R. F. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 487–523 (Univ. Arizona Press, Tucson, 1989).
- Marzari, F., Scholl, H., Murray, C. & Lagerkvist, C. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 725–738 (Univ. Arizona Press, Tucson, 2002).
- Fernandez, J. A. & Ip, W. H. Some dynamical aspects of the accretion of Uranus and Neptune—The exchange of orbital angular momentum with planetesimals. *Icarus* **58**, 109–120 (1984).
- Malhotra, R. The origin of Pluto's peculiar orbit. *Nature* **365**, 819–821 (1993).
- Hahn, J. M. & Malhotra, R. Orbital evolution of planets embedded in a planetesimal disk. *Astron. J.* **117**, 3041–3053 (1999).
- Gomes, R. S., Morbidelli, A. & Levison, H. F. Planetary migration in a planetesimal disk: why did Neptune stop at 30 AU? *Icarus* **170**, 492–507 (2004).
- Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* doi:10.1038/nature03539 (this issue).
- Gomes, R., Tsiganis, K., Morbidelli, A. & Levison, H. F. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* doi:10.1038/nature03676 (this issue).
- Gomes, R. S. Dynamical effects of planetary migration on primordial Trojan-type asteroids. *Astron. J.* **116**, 2590–2597 (1998).
- Michtchenko, T. A., Beaugé, C. & Roig, F. Planetary migration and the effects of mean motion resonances on Jupiter's Trojan asteroids. *Astron. J.* **122**, 3485–3491 (2001).
- Kortenkamp, S. J., Malhotra, R. & Michtchenko, T. Survival of Trojan-type companions of Neptune during primordial planet migration. *Icarus* **167**, 347–359 (2004).
- Levison, H. F., Shoemaker, E. M. & Shoemaker, C. S. Dynamical evolution of Jupiter's Trojan asteroids. *Nature* **385**, 42–44 (1997).
- Milani, A. The Trojan asteroid belt: Proper elements, stability, chaos and families. *Celest. Mech. Dyn. Astron.* **57**, 59–94 (1993).
- Fernández, Y. R., Sheppard, S. S. & Jewitt, D. C. The albedo distribution of Jovian Trojan asteroids. *Astron. J.* **126**, 1563–1574 (2003).
- Barucci, M. A., Cruikshank, D. P., Mottola, S. & Lazzarin, M. in *Asteroids III*

- (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 273–288 (Univ. Arizona Press, Tucson, 2002).
22. Gomes, R. S. The origin of the Kuiper Belt high-inclination population. *Icarus* **161**, 404–418 (2003).
 23. Levison, H. F. & Morbidelli, A. The formation of the Kuiper belt by the outward transport of bodies during Neptune's migration. *Nature* **426**, 419–421 (2003).
 24. Duncan, M. J. & Levison, H. F. A scattered comet disk and the origin of Jupiter family comets. *Science* **276**, 1670–1672 (1997).
 25. Emery, J. P. & Brown, R. H. The surface composition of Trojan asteroids: constraints set by scattering theory. *Icarus* **170**, 131–152 (2004).
 26. Levison, H. F. & Duncan, M. J. From the Kuiper Belt to Jupiter-family comets: The spatial distribution of ecliptic comets. *Icarus* **127**, 13–32 (1997).
 27. Jewitt, D. C., Trujillo, C. A. & Luu, J. X. Population and size distribution of small Jovian Trojan asteroids. *Astron. J.* **120**, 1140–1147 (2000).
 28. Merline, W. J. et al. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 289–314 (Univ. Arizona Press, Tucson, 2002).
 29. Britt, D. T., Yeomans, D., Housen, K. & Consolmagno, G. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 485–500 (Univ. Arizona Press, Tucson, 2002).
 30. Quinn, T. R., Tremaine, S. & Duncan, M. A three million year integration of the earth's orbit. *Astron. J.* **101**, 2287–2305 (1991).

Acknowledgements R.G. is grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico for financial support of his sabbatical year in the OCA observatory in Nice. The work of K.T. was supported by an EC Marie Curie Individual Fellowship. A.M. and H.F.L. thank the CNRS and the NSF for funding the collaboration between the OCA and the SWRI groups. H.F.L. is grateful to NASA's Origins and PG&G programmes.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M. (morby@obs-nice.fr).

LETTERS

Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets

R. Gomes^{1,2}, H. F. Levison^{2,3}, K. Tsiganis² & A. Morbidelli²

The petrology record on the Moon suggests that a cataclysmic spike in the cratering rate occurred ~ 700 million years after the planets formed¹; this event is known as the Late Heavy Bombardment (LHB). Planetary formation theories cannot naturally account for an intense period of planetesimal bombardment so late in Solar System history². Several models have been proposed to explain a late impact spike^{3–6}, but none of them has been set within a self-consistent framework of Solar System evolution. Here we propose that the LHB was triggered by the rapid migration of the giant planets, which occurred after a long quiescent period. During this burst of migration, the planetesimal disk outside the orbits of the planets was destabilized, causing a sudden massive delivery of planetesimals to the inner Solar System. The asteroid belt was also strongly perturbed, with these objects supplying a significant fraction of the LHB impactors in accordance with recent geochemical evidence^{7,8}. Our model not only naturally explains the LHB, but also reproduces the observational constraints of the outer Solar System⁹.

Previous work⁹ explains the current orbital architecture of the planetary system by invoking an initially compact configuration in which Saturn's orbital period was less than twice that of Jupiter. After the dissipation of the gaseous circumsolar nebula, Jupiter's and Saturn's orbits diverged as a result of their interaction with a massive disk of planetesimals, and thus the ratio of their orbital periods, P_S/P_J , increased. When the two planets crossed their mutual 1:2 mean motion resonance (1:2 MMR, that is, $P_S/P_J = 2$) their orbits became eccentric. This abrupt transition temporarily destabilized the giant planets, leading to a short phase of close encounters among Saturn, Uranus and Neptune. As a result of these encounters, and of the interactions of the ice giants with the disk, Uranus and Neptune reached their current heliocentric distances and Jupiter and Saturn evolved to their current orbital eccentricities⁹. The main idea of this Letter is that the same planetary evolution could explain the LHB, provided that Jupiter and Saturn crossed the 1:2 MMR roughly 700 Myr after they formed. Thus, our goal is to determine if there is a generic mechanism that could delay the migration process.

In previous studies^{9–12}, planet migration started immediately because planetesimals were placed close enough to the planets to be violently unstable. Although this type of initial condition was reasonable for the goals of those studies, it is unlikely. Planetesimal-driven migration is probably not important for planet dynamics as long as the gaseous massive solar nebula exists. The initial conditions for the migration simulations should represent the system that existed at the time the nebula dissipated. Thus, the planetesimal disk should contain only those particles that had dynamical lifetimes longer than the lifetime of the solar nebula. In planetary systems like those we adopt from ref. 9, we find that they had to be beyond ~ 15.3 AU (Fig. 1), leading to the initial conditions illustrated in Fig. 2a.

In this configuration, the initial speed of migration would be

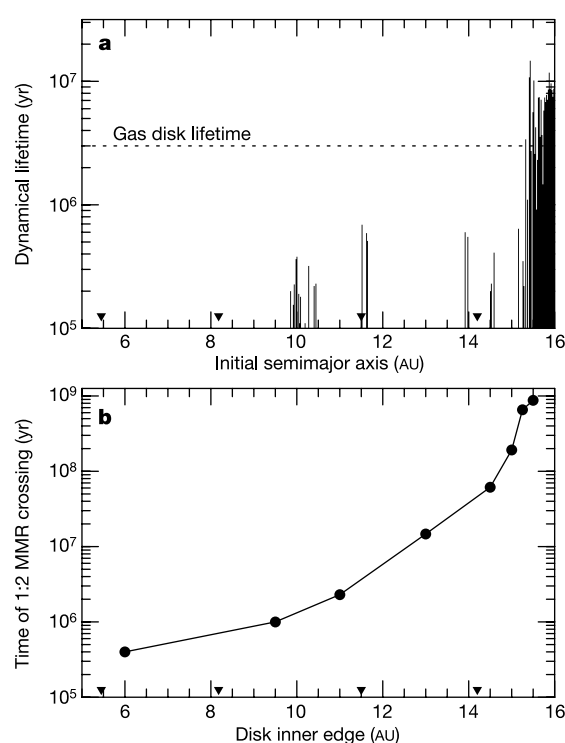


Figure 1 | Disk location and LHB timing. **a**, The histogram reports the average dynamical lifetime of massless test particles placed in a planetary system (shown as triangles) with Jupiter, Saturn and the ice giants on nearly circular, co-planar orbits at 5.45, 8.18, 11.5 and 14.2 AU, respectively. Initially, we placed 10 particles with $e = i = 0$ (where e is eccentricity and i is inclination) and random mean anomaly at each semimajor axis. Stable Trojans of the planets have been removed from this computation. Each vertical bar in the plot represents the average lifetime for those 10 particles. We define ‘dynamical lifetime’ as the time required for a particle to encounter a planet within a Hill radius. A comparison between the histogram and the putative lifetime of the gaseous nebula²⁰ argues that, when the latter dissipated, the inner edge of the planetesimal disk had to be about 1–1.5 AU beyond the outermost ice giant. **b**, Time at which Jupiter and Saturn crossed the 1:2 MMR, as a function of the location of the planetesimal disk’s inner edge, as determined from our first set of migration simulations. In all cases, the disk had a surface density equivalent to $1.9 M_E$ per 1 AU annulus. The outer edge of the disk was varied so that the total mass of the disk was $35 M_E$. The disk was initially very dynamically cold, with $e = 0$ and $i < 0.5^\circ$. A comparison between **a** and **b** shows that a disk that naturally should exist when the nebula dissipated would produce a 1:2 MMR crossing at a time comparable to that of the LHB event.

¹ON/MCT and GEA/OV/UFRJ, Ladeira do Pedro Antonio, 43 Centro 20.080-090, Rio de Janeiro, RJ, Brazil. ²Observatoire de la Côte d’Azur, CNRS, BP 4229, 06304 Nice Cedex 4, France. ³Department of Space Studies, Southwest Research Institute, 1050 Walnut Street, Suite 400, Boulder, Colorado, USA.

dependent on the rate at which disk particles evolve onto planet-crossing orbits. The time at which Jupiter and Saturn cross their 1:2 MMR depends on: (1) their initial distance from the location of the resonance, (2) the surface density of the disk near its inner edge, and (3) the relative location of the inner edge of the disk and the outer ice giant. On the basis of the above arguments, we initially performed a series of eight simulations where the location of the inner edge of the disk was set as the unique free parameter (Fig. 1). As expected, we found a strong correlation between the location of the inner edge and the time of the 1:2 MMR crossing. For disks with inner edges near 15.3 AU (see above), we find crossing times between 192 Myr and 880 Myr (since the beginning of the simulation).

We also performed eight simulations where we varied the initial location of the ice giants by ~ 1 AU, Saturn's location by ~ 0.1 AU, the total mass of the disk by 5 Earth masses ($5M_E$), and its initial dynamical state by pushing the particles' eccentricities up to 0.1 and inclinations up to 3.5° . We found that we can delay the resonant crossing to 1.1 Gyr since the beginning of the simulation, although longer times are clearly possible for more extreme initial conditions. Therefore, we can conclude that the global instability caused by the 1:2 MMR crossing of Jupiter and Saturn could be responsible for the LHB, because the estimated date of the LHB falls in the range of the times that we found.

Figures 2 and 3 show the evolution of one of our runs from the first series of eight. Initially, the giant planets migrated slowly owing to leakage of particles from the disk (Fig. 3a). This phase lasted 880 Myr, at which point Jupiter and Saturn crossed the 1:2 MMR. After the

resonance crossing event, the orbits of the ice giants became unstable and they were scattered into the disk by Saturn. They disrupted the disk and scattered objects all over the Solar System, including the inner regions. The solid curve in Fig. 3b shows the amount of material that struck the Moon as a function of time. A total of 9×10^{21} g struck the Moon after resonance crossing—roughly 50% of this material arrived in the first 3.7 Myr and 90% arrived before 29 Myr. The total mass is consistent with the estimate⁴ of 6×10^{21} g, which was determined from the number and size distribution of lunar basins that formed around the time of the LHB epoch¹. Such an influx spike happened in all our runs. The amount of cometary material delivered to the Earth is $\sim 1.8 \times 10^{23}$ g, which is about 6% of the current ocean mass. This is consistent with upper bounds on the cometary contribution to the Earth's water budget, based on D/H ratio measurement¹³. The average amount of material accreted by the Moon during this spike was $(8.4 \pm 0.3) \times 10^{21}$ g.

The above mass delivery estimate corresponds only to the cometary contribution to the LHB, as the projectiles originated from the external massive, presumably icy, disk. However, our scheme probably also produced an influx of material from the asteroid belt. As Jupiter and Saturn moved from 1:2 MMR towards their current positions, secular resonances (which occur when the orbit of an asteroid processes at the same rate as a planet) swept across the entire belt¹⁴. These resonances can drive asteroids onto orbit with eccentricities and inclinations large enough to allow them to evolve into the inner Solar System and hit the Moon⁴.

We investigated the role of asteroid impactors in our LHB model

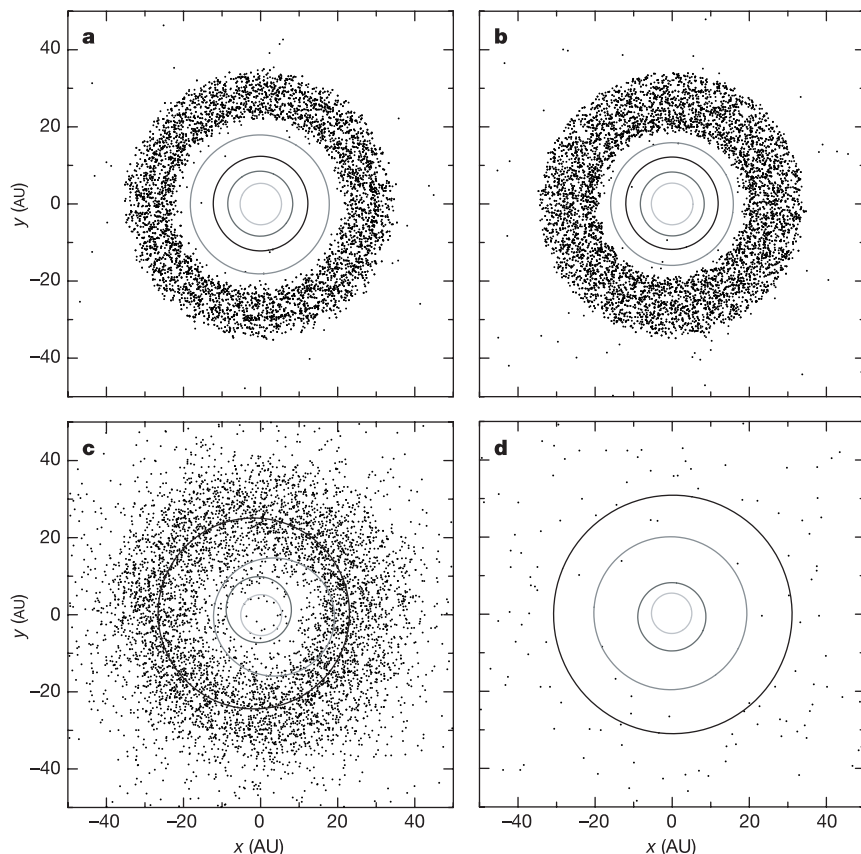


Figure 2 | The planetary orbits and the positions of the disk particles, projected on the initial mean orbital plane. The four panels correspond to four different snapshots taken from our reference simulation. In this run, the four giant planets were initially on nearly circular, co-planar orbits with semimajor axes of 5.45, 8.18, 11.5 and 14.2 AU. The dynamically cold planetesimal disk was $35M_E$, with an inner edge at 15.5 AU and an outer edge

at 34 AU. Each panel represents the state of the planetary system at four different epochs: **a**, the beginning of planetary migration (100 Myr); **b**, just before the beginning of LHB (879 Myr); **c**, just after the LHB has started (882 Myr); and **d**, 200 Myr later, when only 3% of the initial mass of the disk is left and the planets have achieved their final orbits.

by the following numerical integrations. The orbits of an asteroid belt, composed of 1,000 massless particles with semimajor axes between 2.0 and 3.5 AU, were integrated under the gravitational influence of the Sun, Venus, Earth, Mars, Jupiter and Saturn. Because formation models^{15,16} predict that the asteroid belt was partially depleted and dynamically excited well before the LHB, we set the particles' eccentricities between 0 and 0.3 and inclinations between 0° and 30°, but kept the perihelion distances, q , >1.8 AU and aphelion distances, Q , <4 AU. Jupiter and Saturn were forced to migrate at rates that varied from run to run (adopted from ref. 9) by adding a suitably chosen drag-force term to their equations of motion.

We find that objects that reach Earth-crossing orbits follow one of two general paths. Some, referred to as class 1 particles, get trapped in

the periape secular resonance with Saturn (which affects eccentricities) and are driven directly onto Earth-crossing orbits. Other particles, referred to as class 2, stay in the asteroid belt, but are dynamically excited by resonant sweeping onto unstable orbits. These objects slowly leak out of the asteroid belt and can evolve into the inner Solar System. The two classes produce impact spikes with different temporal behaviours. Roughly 50% of class 1 particles arrive in the first 10 Myr, while 90% arrive within ~ 30 Myr. Conversely, the median arrival time for class 2 particles is ~ 50 Myr and 90% arrive within ~ 150 Myr. Class 2 particles dominated in our runs (Fig. 3). However, a preliminary investigation into this issue shows that this result is probably sensitive to the exact evolution of the giant planets and the dynamical state of the asteroid belt. Thus, the best we can conclude is that the impact spike due to asteroids is between these two extremes.

We find that $(3\text{--}8) \times 10^{21}$ g of asteroids hit the Moon during our simulations (Fig. 3). This amount is comparable to the amount of comets. So, our model predicts that the LHB impactors should have been a mixture of comets and asteroids. Unfortunately, we cannot say with any certainty the exact ratio of comets to asteroids in our model because, although the amount of cometary material is fairly well constrained (probably better than a factor of 2), the amount of asteroidal material is not well known (and could be outside the range reported above), because we do not have good estimates of the mass distribution in the asteroid belt before the LHB. It should also be noted that this ratio is probably a function of impactor size, because comets and asteroids probably have different size distributions. This ratio probably also varied with time. Within the first ~ 30 Myr comets dominated according to these simulations, but the last impactors were asteroidal. This is consistent with recent cosmochemical findings suggesting that some of the Moon's basins were formed by asteroids^{7,8}.

Our results support a cataclysmic model for the lunar LHB. Although many aspects of the LHB are not well known¹, our simulations reproduce two of the main characteristics attributed to this episode: (1) the 700 Myr delay between the LHB and terrestrial planet formation, and (2) the overall intensity of lunar impacts. Our model predicts a sharp increase in the impact rate at the beginning of the LHB. Unfortunately, the available lunar data are not yet capable of addressing this prediction.

Our model also has the advantage of supplying impactors that are a mixture of comets and asteroids. Our model predicts that the asteroid belt was depleted by a factor of ~ 10 during the LHB. This depletion does not contradict collisional evolution models^{17,18}. On the contrary, the late secular resonance sweeping could explain why we do not see a large number of asteroid families that were produced during the LHB¹⁹. Our model predicts that the LHB lasted from between ~ 10 Myr and ~ 150 Myr. Correspondingly, the drop-off in impact rates could be quite fast (with 50% of the impacts occurring in the first 3.7 Myr and 90% in 29 Myr) or moderately slow (with 50% of the impacts occurring in the first 50 Myr and 90% in 150 Myr). We are unable to pinpoint more exact values because the duration and the drop-off of the LHB depends on the relative contributions of class 1 asteroids, class 2 asteroids, and comets, which in turn are very sensitive to the pre-LHB orbital structure of the asteroid belt.

Most importantly, our scheme for the LHB is the result of a generic migration-delaying mechanism, followed by an instability, which is itself induced by a deterministic mechanism of orbital excitation of the planets⁹. This revised planetary migration scheme naturally accounts for the currently observed planetary orbits⁹, the LHB, the present orbital distribution of the main-belt asteroids and the origin of Jupiter's Trojans¹⁹.

Received 6 December 2004; accepted 18 April 2005.

1. Hartmann, W. K., Ryder, G., Dones, L. & Grinspoon, D. in *Origin of the Earth and Moon* (eds Canup, R. & Righter, K.) 493–512 (Univ. Arizona Press, Tucson, 2000).

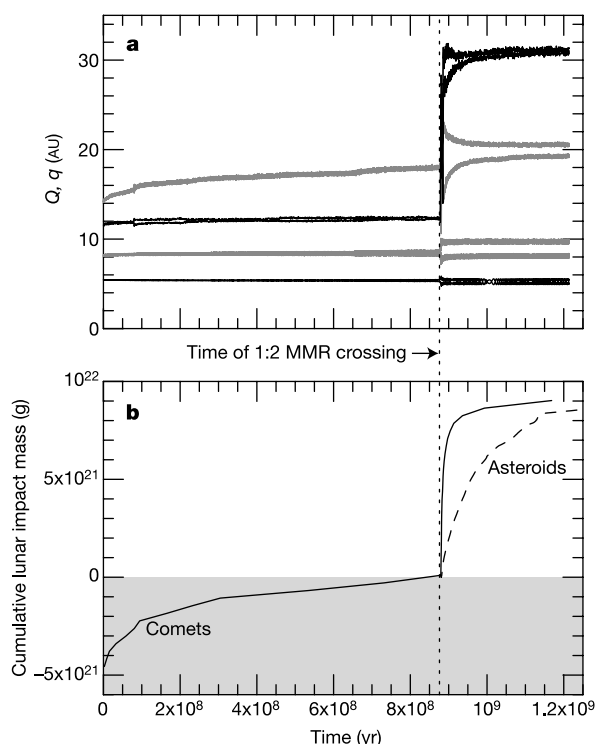


Figure 3 | Planetary migration and the associated mass flux towards the inner Solar System from a representative simulation. **a**, The evolution of the four giant planets. Each planet is represented by a pair of curves—the top and bottom curves are the aphelion and perihelion distances, Q and q , respectively. Jupiter and Saturn cross the 1:2 MMR at 880 Myr. The subsequent interaction between the planets and the disk led to the current planetary configuration as shown in ref. 9. **b**, The cumulative mass of comets (solid curve) and asteroids (dashed curve) accreted by the Moon. We have offset the comet curve so that the value is zero at the time of 1:2 MMR crossing. Thus, $\sim 5 \times 10^{21}$ g of comets was accreted before resonant crossing and 9×10^{21} g of cometary material would have struck the Moon during the LHB. Although the terrestrial planets were not included in our cometary simulations, we estimated the amount of material accreted by the Moon directly from the mass of the planetesimal disk by combining the particles' dynamical evolution with the analytic expressions in ref. 21. The impact velocity of these objects ranged from 10 to 36 km s^{-1} with an average of 21 km s^{-1} . Estimating the asteroidal flux first requires a determination of the mass of the asteroid belt before resonant crossing. This value was determined by first combining the percentage of asteroids remaining in the belt at the end of a simulation ($\sim 10\%$, very sensitive to planet migration rate and initial asteroid distribution) with estimates of the current mass of the belt to determine the initial asteroid belt mass ($\sim 5 \times 10^{-3} M_{\oplus}$). The flux was then again determined by combining the particles' dynamical evolution with the analytic expressions in ref. 21. The dashed curve shows a simulation where class 2 particles dominate. The average asteroidal impact velocity is 25 km s^{-1} .

2. Morbidelli, A., Petit, J.-M., Gladman, B. & Chambers, J. A plausible cause of the Late Heavy Bombardment. *Meteorit. Planet. Sci.* **36**, 371–380 (2001).
3. Zappala, V., Cellino, A., Gladman, B. J., Manley, S. & Migliorini, F. Asteroid showers on Earth after family break-up events. *Icarus* **134**, 176–179 (1998).
4. Levison, H. F. *et al.* Could the lunar “Late Heavy Bombardment” have been triggered by the formation of Uranus and Neptune? *Icarus* **151**, 286–306 (2001).
5. Chambers, J. E. & Lissauer, J. J. A new dynamical model for the lunar Late Heavy Bombardment. *Lunar Planet. Sci. Conf. XXXIII*, abstr. 1093 (2002).
6. Levison, H. F., Thommes, E. W., Duncan, M. J., Dones, L. A. in *Debris Disks and the Formation of Planets: A Symposium in Memory of Fred Gillett (11–13 April 2002, Tucson, Arizona)* (eds Caroff, L., Moon, L. J., Backman, D. & Praton, E.) 152–167 (ASP Conf. Ser. 324, Astronomical Society of the Pacific, San Francisco, 2005).
7. Kring, D. A. & Cohen, B. A. Cataclysmic bombardment throughout the inner Solar System 3.9–4.0 Ga. *J. Geophys. Res. Planets* **107**(E2), 4–10 (2002).
8. Tagle, R. LL-ordinary chondrite impact on the Moon: Results from the 3.9 Ga impact melt at the landing site of Apollo 17. *Lunar Planet. Sci. Conf. XXXVI*, abstr. 2008 (2005).
9. Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* doi:10.1038/nature03539 (this issue).
10. Fernandez, J. A. & Ip, W.-H. Some dynamical aspects of the accretion of Uranus and Neptune—The exchange of orbital angular momentum with planetesimals. *Icarus* **58**, 109–120 (1984).
11. Hahn, J. M. & Malhotra, R. Orbital evolution of planets embedded in a planetesimal disk. *Astron. J.* **117**, 3041–3053 (1999).
12. Gomes, R. S., Morbidelli, A. & Levison, H. F. Planetary migration in a planetesimal disk: Why did Neptune stop at 30 AU? *Icarus* **170**, 492–507 (2004).
13. Morbidelli, A. *et al.* Source regions and timescales for the delivery of water to Earth. *Meteorit. Planet. Sci.* **35**, 1309–1320 (2000).
14. Gomes, R. S. Dynamical effects of planetary migration on the primordial asteroid belt. *Astron. J.* **114**, 396–401 (1997).
15. Wetherill, G. W. An alternative model for the formation of the asteroids. *Icarus* **100**, 307–325 (1992).
16. Petit, J., Morbidelli, A. & Chambers, J. The primordial excitation and clearing of the asteroid belt. *Icarus* **153**, 338–347 (2001).
17. Davis, D. R., Ryan, E. V. & Farinella, P. Asteroid collisional evolution: results from current scaling algorithms. *Planet. Space Sci.* **42**, 599–610 (1994).
18. Bottke, W. *et al.* The fossilized size distribution of the main asteroid belt. *Icarus* **175**(1), 111–140 (2005).
19. Morbidelli, A., Levison, H. F., Tsiganis, K. & Gomes, R. Chaotic capture of Jupiter’s Trojan asteroids in the early Solar System. *Nature* doi:10.1038/nature03540 (this issue).
20. Haisch, K. E., Lada, E. A. & Lada, C. J. Disk frequencies and lifetimes in young clusters. *Astrophys. J.* **553**, L153–L156 (2001).
21. Wetherill, G. W. Collisions in the asteroid belt. *J. Geophys. Res.* **72**, 2429–2444 (1967).

Acknowledgements R.G. thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico for support for his sabbatical year in the OCA observatory in Nice. K.T. was supported by an EC Marie Curie Individual Fellowship. A.M. and H.F.L. thank the CNRS and the NSF for funding collaboration between the OCA and the SWRI groups. H.F.L. was supported by NASA’s Origins and PG&G programmes.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M. (morby@obs-nice.fr).

LETTERS

Quantum interference during high-order harmonic generation from aligned molecules

Tsuneto Kanai, Shinichirou Minemoto & Hirofumi Sakai

High-order harmonic generation (HHG) from atoms and molecules offers potential application as a coherent ultrashort radiation source in the extreme ultraviolet and soft X-ray regions^{1–3}. In the three-step model^{4–6} of HHG, an electron tunnels out from the atom and may recombine with the parent ion (emitting a high-energy photon) after undergoing laser-driven motion in the continuum. Aligned molecules^{7–11} can be used to study quantum phenomena in HHG associated with molecular symmetries; in particular, simultaneous observations of both ion yields and harmonic signals under the same conditions serve to disentangle the contributions from the ionization and recombination processes. Here we report evidence for quantum interference of electron de Broglie waves^{12–14} in the recombination process of HHG from aligned CO₂ molecules. The interference takes place within a single molecule and within one optical cycle. Characteristic modulation patterns of the harmonic signals measured as a function of the pump–probe delay are explained with simple formulae determined by the valence orbital of the molecules. We propose that simultaneous observations of both ion yields and harmonic signals can serve as a new route to probe the instantaneous structure of molecular systems.

HHG from atoms is well studied and understood owing to their central symmetry. On the other hand, molecules are not isotropic systems, and random alignment of sample molecules has prevented researchers from studying the detailed physics of HHG from molecules. But recently developed molecular alignment^{7–10} and orientation¹¹ techniques have now begun to reveal the physics of HHG from molecules, stemming from their various symmetries.

HHG from atoms is well explained by the three-step model^{4–6}. First, an electron tunnels through the potential barrier modified by the atomic potential and the intense laser field, and appears in the continuum (step 1). The free electron is then driven by the laser field and has a probability of returning to the parent ion after the field reverses its direction (step 2). A high-energy photon is emitted if the recollision with the parent ion leads to recombination (step 3). This simple picture explains not only HHG but also other interesting phenomena, such as non-sequential ionization¹⁵ and above-threshold-ionization (ATI)¹⁶. A maximum kinetic energy $K_{\text{ret}} = 3.17 U_p$ for the returning electron is predicted, which leads to a well-known cut-off law for the photon energy, $\omega_{\text{HH}} \leq I_p + 3.17 U_p$. Here I_p is the ionization potential of the atom, and U_p is the ponderomotive potential in the laser field. As long as a sample of randomly aligned molecules is employed, high-order harmonics from molecules have shown generation characteristics similar to those from atoms^{17–19}. This means that the basic idea of the three-step model can also be applied to HHG from molecules.

However, with a sample of aligned molecules, there is an opportunity to investigate the contributions from each step of the three-step model for HHG from molecules. As for step 1 of HHG (ionization), recent experimental⁸ and theoretical^{20,21} studies reveal

that ionization rates depend on the molecular orientation, and this dependence stems from the violation of the central symmetry of the valence orbital. If the other two steps do not counteract the effects of step 1, HHG is enhanced when ionization is enhanced. This applies to the recent observations of high-order harmonics from non-adiabatically aligned N₂ (refs 22, 23) via the pump (for alignment) and probe (for HHG) technique.

If steps 2 and 3 (here we call them ‘recombination’ *en bloc*) play a crucial role, HHG does not necessarily correlate with ionization any more. In fact, the main phenomenon expected to be observed in aligned molecules is quantum interference in the recombination process, as pointed out in recent theoretical studies^{12–14}. Here we report clear evidence of quantum interference in the recombination process using aligned CO₂ molecules. The observation of quantum interference has been made possible by the simultaneous observations of high-order harmonics and ion yields under the same experimental conditions (see Methods). In order to explain the observed interference effect, we successfully apply a model of two point emitters, which was originally proposed for a diatomic molecule^{12,13}, to a triatomic molecule. On the basis of our way of applying the two-point-emitter model, a CO₂ molecule can be regarded as an elongated O₂ molecule with the same symmetry (π_g).

Figure 1 shows the evolution of the ion yields (Fig. 1a) and the 23rd

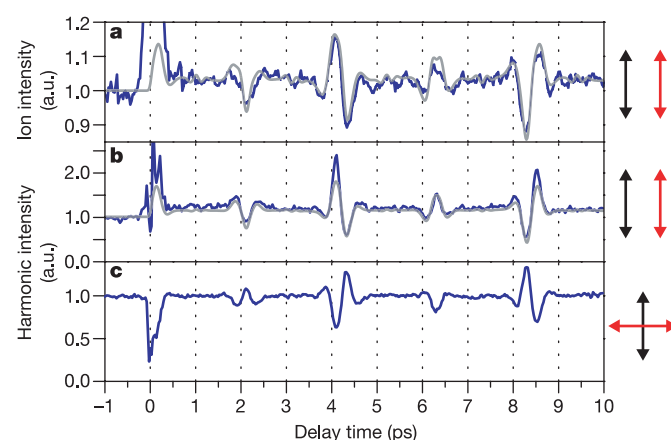


Figure 1 | The time evolution of the ion yield and the 23rd harmonic intensity from N₂ molecules. **a–c**, The ion yield dominated by N₂⁺ (a) and the 23rd harmonics from N₂ molecules (b, c) as a function of pump–probe delay. The polarizations of the pump (black arrow) and the probe (red arrow) pulses are parallel (a, b) or perpendicular (c) to each other. Note that the rotational period T_{rot} of a N₂ molecule is 8.4 ps. The results of theoretical calculations are shown by grey curves. The rotational temperature of the N₂ molecules is assumed to be 80 K, and the highest degree of alignment $\langle \cos^2\theta \rangle$ is estimated to be 0.62 at $T_{\text{rot}}/2$. a.u., arbitrary units.

harmonics (Fig. 1b, c) from N_2 molecules as a function of the delay (τ) between the pump and the probe pulses. The polarizations of the pump and the probe pulses are parallel (in Fig. 1a, b) or perpendicular (in Fig. 1c) to each other. In both the ion yields and the harmonics, the intensities modulate at a period of ~ 2 ps, which is a quarter of the rotational period T_{rot} (~ 8.4 ps) of neutral N_2 molecules. The degree of modulation is small at $\tau \approx T_{\text{rot}}/4$ and $3T_{\text{rot}}/4$, and it is large at $\tau \approx T_{\text{rot}}/2$ and T_{rot} , which is explained by the nuclear spin statistics²⁴. The modulations at every $T_{\text{rot}}/4$ are characteristic of molecules whose highest occupied molecular orbital (HOMO) has σ_g symmetry. When the polarizations are parallel to each other, the harmonic intensity (Fig. 1b) modulates in phase with the ion yields (Fig. 1a). As the ion yields reflect the degree of alignment along the polarization of the probe pulses⁸, the correlation between the ionization and the HHG shows that N_2 molecules aligned along the laser field efficiently generate high-order harmonics, while the anti-aligned ones suppress them.

We can express the angular dependence of the ion yield or the harmonic intensity $\langle I(\theta) \rangle$ as

$$\begin{aligned} \langle I(\theta) \rangle &\propto \langle P_0(\cos\theta) + a_2P_2(\cos\theta) + a_4P_4(\cos\theta) \rangle \\ &= (1 - a_2/2 + 3a_4/8) + (3a_2/2 + 5a_4/8)\langle \cos^2\theta \rangle \\ &\quad - (35a_4/32)\langle \sin^2\theta \rangle \end{aligned} \quad (1)$$

where the orientation angle θ is the angle between the molecular axis and the polarization direction of the probe pulses, the $P_n(\cos\theta)$ terms are the n th Legendre polynomials and the a_n terms are constants (see Methods). The average of $\cos^2\theta$, that is, $\langle \cos^2\theta \rangle$ in equation (1), represents the degree of molecular alignment. The results of theoretical calculations are also included in Fig. 1. We see satisfactory agreement between the experiments and the calculations, which ensures the validity of equation (1). In the case of σ_g -type molecules, the contribution of the third term in equation (1) is small compared to that of the second term (see Methods). Figure 1c shows that the phase of the modulation is inverted when the polarization of the probe pulse is rotated by 90° . This observation is explained by the dominant second term, because $\langle \cos^2(\theta - \pi/2) \rangle = \langle \sin^2\theta \rangle = 1 - \langle \cos^2\theta \rangle$.

As an example of π_g -type molecules, we first investigate O_2 molecules. The experimental results are shown in Fig. 2, and are in good agreement with the results of theoretical calculations. For molecules whose HOMOs have π_g symmetry, the contribution of the third term in equation (1) becomes comparable to that of the second term (see Methods). The modulations of the ion yield and the harmonic intensity for O_2 molecules (Fig. 2) are quite different from those for N_2 (Fig. 1). Considering $T_{\text{rot}} = 11.6$ ps for O_2 , we can see a modulation at every one-eighth of T_{rot} , which is characteristic of molecules whose HOMOs have anti-bonding π_g symmetry. As in the case of N_2 molecules, the harmonic intensity (Fig. 2b) modulates in phase with the ion intensity (Fig. 2a). Furthermore, when the polarizations are perpendicular to each other (Fig. 2c), the harmonic intensity modulates out of phase with that in the parallel case (Fig. 2b). This modulation inversion can be understood by a simple analysis with equation (1) (see Methods). Note that the second term in equation (1), which naturally emerges in the expansion, represents the violation of four-fold symmetry. This effect on the ionization rate is taken into account in the molecular ADK model²¹. The present observations suggest that the harmonic intensity from O_2 molecules takes its maximum when the orientation angle θ is deviated from 45° .

In order to further investigate our observation that the symmetry of the HOMO plays a crucial role in HHG, CO_2 molecules are also adopted as a nonlinear medium. The HOMO of a CO_2 molecule, like that of an O_2 molecule, has anti-bonding π_g symmetry and is dominated by the two p orbitals of the two O atoms. The difference between O_2 and CO_2 molecules is in the distance R between the two O atoms; R of CO_2 (0.232 nm) is almost twice as long as R of O_2

(0.121 nm). Because of this difference in R , we can investigate the influence of R on the HHG process. Figure 3, together with the results of theoretical calculations, shows the delay dependence of the ion yield and the 23rd harmonic intensity from CO_2 molecules. Reasonable agreement between the experiments and the calculations can be confirmed again. The discrepancy at larger delay times in Fig. 3b might be due to some changes in the experimental conditions during the measurements. Considering $T_{\text{rot}} = 42.7$ ps for linear CO_2 molecules, the modulation period (~ 5.4 ps) corresponds to $T_{\text{rot}}/8$ and this fractional revival (of the quantum mechanical state of molecular alignment) is observed, as in the case of O_2 molecules. The modulation period of $T_{\text{rot}}/8$ ensures that linear CO_2 molecules rather than bent CO_2 molecules dominantly contribute to HHG. As in the case of N_2 and O_2 molecules, $\langle I(\theta) \rangle$ in the parallel case (Fig. 3b) and $\langle I(\theta) \rangle$ in the perpendicular case (Fig. 3c) show the inverted modulations, which can be explained by equation (1) with different constants. However, the ion yield in the parallel case (Fig. 3a) and the harmonic intensity in the parallel case (Fig. 3b) also show the inverted modulations as opposed to the cases for N_2 and O_2 molecules (see Methods, and compare the signs of the coefficients in equation (1) for CO_2 to those for N_2 and O_2).

These observations indicate that HHG is enhanced when ionization is suppressed. This means that the recombination probabilities at a certain orientation angle overcome the ionization process (step 1). Our observations of anticorrelation between ion yields and harmonic signals for the 19th to the 29th harmonic cannot be explained by a three-centre effective potential model as proposed for the HHG in triatomic molecules²⁵. This may be due to the fact that the harmonic orders (9th to 17th) investigated in ref. 25 are different from those observed here (19th to 29th). As explained below, the modulation inversions observed in Fig. 3a and b are, to our knowledge, the first clear evidence of quantum interference in the recombination process.

The basic idea of the molecular ADK model²¹, which was originally developed for diatomic molecules, is that the HOMO determines the angular dependence of ionization rates. As mentioned above, the HOMO of a CO_2 molecule is dominated by the two p orbitals of the two O atoms. Therefore, as a natural extension of the molecular ADK model, the ionization process of a CO_2 molecule can be considered to be equivalent to that of an elongated O_2 molecule whose HOMO has the same anti-bonding π_g symmetry as that of CO_2 . This extended

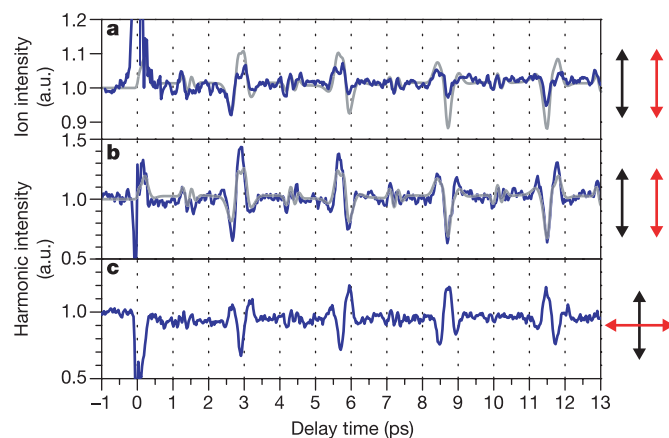


Figure 2 | The time evolution of the ion yield and the 23rd harmonic intensity from O_2 molecules. **a–c**, The ion yield dominated by O_2^+ (**a**) and the 23rd harmonics from O_2 molecules (**b**, **c**) as a function of pump–probe delay. The experimental conditions are identical with those of Fig. 1. Note that the rotational period T_{rot} of an O_2 molecule is 11.6 ps. The results of theoretical calculations are shown by grey curves. The rotational temperature of the O_2 molecules is assumed to be 80 K, and the highest degree of alignment $\langle \cos^2\theta \rangle$ is estimated to be 0.62 at $T_{\text{rot}}/4$.

molecular ADK model predicts that the maximum ionization rate of CO₂ should be achieved at $\theta \approx 30^\circ$. Our counterintuitive observations in Fig. 3 suggest the crucial role of the recombination process in HHG from CO₂ molecules.

Recently, Lein *et al.* pointed out that, according to theory, for any phenomenon including electron recollision such as HHG and ATI, interference can take place in diatomic molecules^{12–14}. The interference leads to a peculiar orientation dependence of recombination probabilities. Their detailed calculations can be interpreted by a simple model of two point emitters situated at the positions of the nuclei. Here, we apply their model to a triatomic molecule in the following way. As in the case of the ionization process, a CO₂ molecule can be regarded as an elongated diatomic molecule where the point emitters are located in the two O nuclei. The conditions for interference to take place in a diatomic molecule whose HOMO has anti-bonding symmetry (for example, O₂) are

$$R \cos \theta = n\lambda \text{ (destructive)} \quad (2)$$

$$R \cos \theta = (n - 1/2)\lambda \text{ (constructive)} \quad (3)$$

where λ is the de Broglie wavelength of a free electron, R is the internuclear distance of the diatomic molecule, and n is a positive integer^{12,14}.

The model of two point emitters applied here is illustrated in Fig. 3d. In equations (2) and (3), only $n = 1$ can be relevant to harmonics of relatively low energy (≤ 60 eV) from small neutral molecules ($R \leq 0.2$ nm and $I_p \leq 20$ eV). Our detailed quantum mechanical calculations show that the destructive interference in CO₂ strongly suppresses the recombination probability at $\theta \approx 30^\circ$ by orders of magnitude, though the ionization of CO₂ is maximized at $\theta \approx 30^\circ$, which is in turn determined by the molecular ADK theory²¹. It should be noted that different interference effects between the electron trajectories are included in our calculations, and that the complicated evolution of the molecular alignment is taken into account, though we use $\theta \approx 30^\circ$ as the distribution peak to simplify the discussion. With an effective wave vector, for which I_p is set to be zero¹³, equation (2) predicts that, at $\theta \approx 30^\circ$, the destructive interference takes place at the 23rd harmonics, which is consistent with our observation shown in Fig. 3. Here, we note that the destructive interference is also reflected in the downward shift of the base line of

the 23rd harmonics in CO₂ at $\tau > 0$ from that at $\tau < 0$ (Fig. 3b). A CO₂ molecule can be thought of as a system where the distance between the two point emitters is appropriately long to induce such a strong interference effect. Appropriately elongated diatomic molecules whose HOMOs have anti-bonding symmetry should show modulation patterns very similar to those in Fig. 3. We note that the above conditions (equations (2) and (3)) for the destructive and constructive interference are interchanged for molecules whose HOMOs have bonding symmetry (for example, N₂) according to their relative intramolecular phase^{12,14}.

Equations (2) and (3) tell us that the interference also depends on the wavelength of the free electron, that is, the order of the harmonics. Therefore, the modulation in the delay dependence of HHG should depend on the harmonic order. The harmonic order dependence of the interference effect is investigated both for CO₂ and for N₂ for the purpose of comparison. With the polarizations parallel to each other, the harmonic spectra are measured at fixed time delays when ionization is enhanced ($\tau = 4.1$ and 20.9 ps for N₂ and CO₂, respectively) and suppressed ($\tau = 4.2$ and 21.9 ps). Figure 4 shows the integrated signals of high-order harmonics thus measured for N₂ (Fig. 4a) and CO₂ (Fig. 4b). The signals are normalized to those from randomly aligned molecules. For CO₂ molecules (Fig. 4b), most orders of measured harmonic signals including the 23rd shown in Fig. 3 anticorrelate with ion signals; harmonic signals are suppressed when ionization is enhanced at $\tau = 20.9$ ps, while they are enhanced at $\tau = 21.9$ ps. In contrast to the results for CO₂ molecules, most orders of harmonic signals from aligned or anti-aligned N₂ molecules correlate with ion signals as observed in Fig. 1. Still, the destructive interference effect (close to anticorrelation) can be seen at a higher order, the 31st. The results obtained from detailed quantum mechanical calculations of the recombination process are also shown in Fig. 4. The anticorrelation for CO₂ as well as the correlation for N₂ are successfully reproduced. Furthermore, the agreement between the experiments and the calculations is reasonably good, and supports our interpretations of quantum interference. Better agreement should be achieved by including higher order Legendre polynomials in equation (1) (especially for CO₂). As pointed out in ref. 13, the interference observed here may be interpreted as a microscopic version of the 200-yr-old Young's two-slit experiment²⁶.

In the case of O₂ molecules, on the other hand, the interference

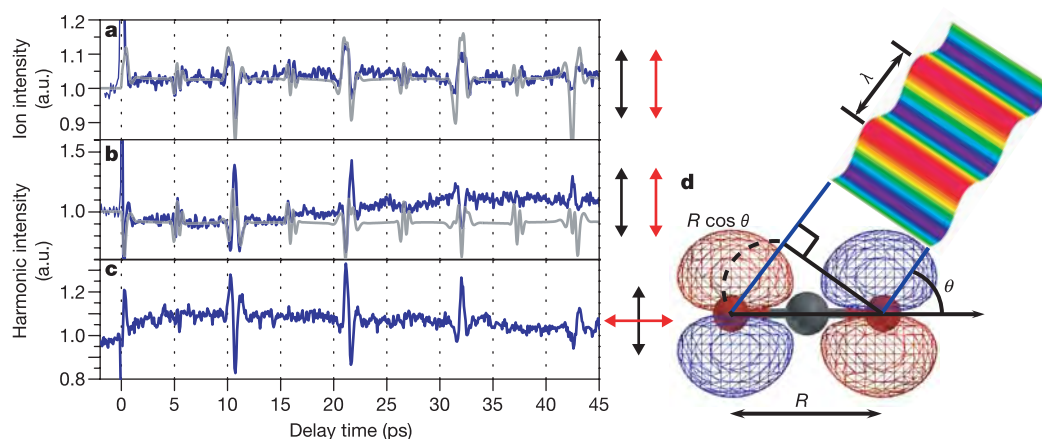


Figure 3 | The time evolution of the ion yield and the 23rd harmonic intensity from CO₂ molecules, and an illustration of the model of two point emitters. **a–c**, The ion yield dominated by CO₂⁺ (**a**) and the 23rd harmonics from CO₂ molecules (**b**, **c**) as a function of pump–probe delay. The experimental conditions are identical with those of Fig. 1. Note that the rotational period T_{rot} of a CO₂ molecule is 42.7 ps. The results of theoretical calculations are shown by grey curves. The rotational temperature of CO₂ molecules is assumed to be 40 K, and the highest degree of alignment ($\cos^2\theta$) is estimated to be 0.70 at $3T_{\text{rot}}/4$. **d**, A CO₂ molecule can be regarded as an

appropriately elongated diatomic molecule. Two point emitters are located in two O nuclei depicted as red spheres. λ is the de Broglie wavelength of a free electron. θ is the orientation angle, that is, the angle between the molecular axis and the polarization direction of the probe pulse. R and $R \cos \theta$ are the distance between two O atoms and its projection, respectively. In the recombination process, the destructive or constructive interference takes place depending on the conditions (equations (2) and (3)) given in the text.

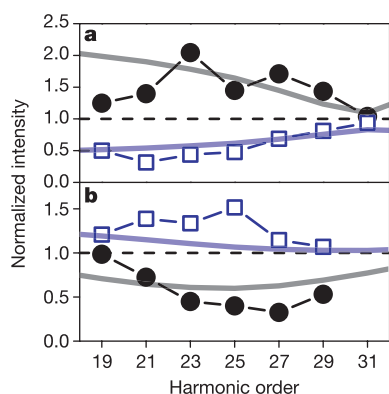


Figure 4 | The integrated signals of high-order harmonics measured as a function of the harmonic order. Results are shown for N_2 (a) and CO_2 (b). The signals are normalized to those from randomly aligned molecules. With the polarizations parallel to each other, the harmonic spectra are measured at fixed time delays when ionization is enhanced (filled circles, $\tau = 4.1$ and 20.9 ps for N_2 and CO_2 , respectively) and suppressed (open squares, $\tau = 4.2$ and 21.9 ps). The results of theoretical calculations are shown by the black and the blue curves, which correspond to the time delays when ionization is enhanced and suppressed, respectively.

effect cannot be observed under the present experimental conditions, because the cut-off is longer than the appropriate wavelength regions of harmonics for which the interference effect can take place. With probe pulses having shorter pulse widths²⁷, even for O_2 molecules, one might be able to observe the interference effect (anticorrelation) for the higher orders of harmonics, as suggested in the results for N_2 (Fig. 4a).

Here we propose a potential application of the present technique to probe the instantaneous structure of the molecular systems. The procedures of this new probe should be as follows. (1) Prepare a sample of highly aligned molecules, so that we can regard $\cos\theta$ as unity in equations (2) and (3) (anti-bonding case) or the interchanged conditions (bonding case). (2) Observe the angular dependence of ion yields ($I_{ion}(\theta)$) and harmonic signals ($I_{HH}(\theta)$) of an appropriate harmonic order (for which destructive interference can be expected) from aligned molecules by rotating the polarization direction of the probe pulse. (3) Plot a graph of the ratio $I_{HH}(\theta)/I_{ion}(\theta)$ against θ , which should give the angular dependence of the recombination process $I_{recom}(\theta)$ (note that $I_{HH}(\theta) \propto I_{ion}(\theta) \times I_{recom}(\theta)$). (4) Find the angle θ at which a dip occurs in the graph. (5) Then the instantaneous bond length R can be determined by the condition for destructive interference (equation (2) in the case of anti-bonding). The accuracy of the measurement could be less than one optical cycle if we employ few-cycle pulses as probe pulses.

As demonstrated in the present studies, a sample of aligned molecules serves as an ideal quantum system to investigate the quantum phenomena associated with molecular symmetries as well as the contributions from each step in the three-step model for HHG. One of the next challenges is to generate high-order harmonics from oriented molecules. With a sample of oriented molecules, we can expect the generation of even-order harmonics because of the breaking of the inversion symmetry.

METHODS

Experimental. We prepared the femtosecond pump (for alignment) and probe (for HHG) pulses with a Michelson-type interferometer. An output from a Ti:sapphire based chirped pulse amplification system (Spectra-Physics, Super-Spitfire) with a pulse width of ~ 50 fs and a centre wavelength of ~ 800 nm is split into two pulses. The first pulse is used as a pump to create rotational wavepackets and induce non-adiabatic molecular alignment. The second pulse is delayed by a computer controlled translation stage, and is used as a probe to generate high-order harmonics. The two pulses are focused with a lens ($f = 300$ mm) into a

supersonic gas jet in the vacuum chamber. The intensity of the pump pulse is $\sim 6 \times 10^{13} \text{ W cm}^{-2}$ and that of the probe is $\sim 2 \times 10^{14} \text{ W cm}^{-2}$, which is below the saturation intensity. The generated harmonics are spectrally resolved by a 1-m grazing incidence monochromator (McPherson, Model 248/310G) with a Pt-coated 300-grooves mm^{-1} grating and detected by an electron multiplier. The ions produced through multiphoton ionization are detected by a cylindrical ion collector prepared downstream of the gas jet²⁸. The harmonic and ion signals are accumulated by a digital oscilloscope and transferred to and processed in a personal computer.

Angular dependence of ion yield and harmonic intensity. In general, the angular dependence of physical phenomena resulting from linear molecules with axial symmetry can be expressed by the expansion of orthonormal functions such as Legendre polynomials, $P_n(\cos\theta)$ s. In the case of linear molecules with $D_{\infty h}$ symmetry, only the even order $P_n(\cos\theta)$ terms should be included in the expansion, as shown in equation (1). Higher accuracy of the expansion can be achieved by including higher order Legendre polynomials. The coefficients for ion yields and those for harmonic signals are determined by the molecular ADK theory²¹ and by the detailed quantum mechanical calculations for the recombination process⁶, respectively. In order to calculate the time evolution of expectation values of Legendre polynomials in equation (1), we have developed a calculation procedure to integrate time-dependent Schrödinger equations. The details of the theoretical calculations will be reported elsewhere. The coefficients in equation (1) employed to fit the time evolution of the ion yield are $a_2 = 0.39$, $a_4 = -0.21$ for N_2 (Fig. 1a), $a_2 = 0.34$, $a_4 = -0.36$ for O_2 (Fig. 2a), and $a_2 = 0.41$, $a_4 = -0.38$ for CO_2 (Fig. 3a). Those employed to fit the time evolution of the 23rd harmonic intensity are $a_2 = 2.05$, $a_4 = -0.52$ for N_2 (Fig. 1b), $a_2 = 0.83$, $a_4 = -1.17$ for O_2 (Fig. 2b), and $a_2 = -0.11$, $a_4 = -1.44$ for CO_2 (Fig. 3b).

The modulation inversion observed in Fig. 2b and c can be explained as follows. Because $\langle \cos^2(\theta - \pi/2) \rangle = 1 - \langle \cos^2\theta \rangle$ and $\langle \sin^2 2(\theta - \pi/2) \rangle = \langle \sin^2 2\theta \rangle$, the harmonic intensity for the perpendicular configuration is $\langle I(\theta) \rangle \sim (1 - a_2/2 + 3a_4/8) + (3a_2/2 + 5a_4/8)(1 - \langle \cos^2\theta \rangle) - (35a_4/32)\langle \sin^2 2\theta \rangle$. The modulation is reversed by $(3a_2/2 + 5a_4/8)(1 - 2\langle \cos^2\theta \rangle)$ from that in the parallel configuration. Strictly speaking, in the replacement of θ with $\theta - \pi/2$ account should be taken of the distribution of molecules and the transformation of the volume element, though the resulting expression becomes more complicated.

We notice that in the case of CO_2 the harmonic signals at $T_{rot}/8$ modulate in phase with the ion signals, though this is difficult to recognize in Fig. 3. This behaviour is also successfully reproduced by theoretical calculations. The different behaviours of HHG at $T_{rot}/8$ and $T_{rot}/4$ correspond to the different quantum mechanical states of alignment at the two fractional revivals.

Received 17 August 2004; accepted 1 April 2005.

- Hentschel, M. *et al.* Attosecond metrology. *Nature* **414**, 509–513 (2001).
- Drescher, M. *et al.* Time-resolved atomic inner-shell spectroscopy. *Nature* **419**, 803–807 (2002).
- Chang, Z., Rundquist, A., Wang, H., Murnane, M. M. & Kapteyn, H. C. Generation of coherent soft X rays at 2.7 nm using high harmonics. *Phys. Rev. Lett.* **79**, 2967–2970 (1997).
- Corkum, P. B. Plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Krause, J. L., Schafer, K. J. & Kulander, K. C. High-order harmonic generation from atoms and ions in the high intensity regime. *Phys. Rev. Lett.* **68**, 3535–3538 (1992).
- Lewenstein, M. *et al.* Theory of high-harmonic generation by low-frequency laser fields. *Phys. Rev. A* **49**, 2117–2132 (1994).
- Rosca-Pruna, F. & Vrakking, M. J. J. Experimental observation of revival structures in picosecond laser-induced alignment of I_2 . *Phys. Rev. Lett.* **87**, 153902 (2001).
- Litvinyuk, I. V. *et al.* Alignment-dependent strong field ionization of molecules. *Phys. Rev. Lett.* **90**, 233003 (2003).
- Sakai, H. *et al.* Controlling the alignment of neutral molecules by a strong laser field. *J. Chem. Phys.* **110**, 10235–10238 (1999).
- Velotta, R., Hay, H., Mason, M. B., Castillejo, M. & Marangos, J. P. High-order harmonic generation in aligned molecules. *Phys. Rev. Lett.* **87**, 183901 (2001).
- Sakai, H., Minemoto, S., Nanjo, H., Tanji, H. & Suzuki, T. Controlling the orientation of polar molecules with combined electrostatic and pulsed, nonresonant laser fields. *Phys. Rev. Lett.* **90**, 083001 (2003).
- Lein, M., Hay, N., Velotta, R., Marangos, J. P. & Knight, P. L. Role of the intramolecular phase in high-harmonic generation. *Phys. Rev. Lett.* **88**, 183903 (2002).
- Lein, M., Hay, N., Velotta, R., Marangos, J. P. & Knight, P. L. Interference effects in high-order harmonic generation with molecules. *Phys. Rev. A* **66**, 023805 (2002).
- Lein, M., Corso, P. P., Marangos, J. P. & Knight, P. L. Orientation dependence of high-order harmonic generation in molecules. *Phys. Rev. A* **67**, 023819 (2003).

15. Sakai, H. *et al.* Nonsequential double ionization of D₂ molecules with intense 20-fs pulses. *Phys. Rev. A* **67**, 063404 (2003).
16. Lein, M., Marangos, J. P. & Knight, P. L. Electron diffraction in above-threshold ionization of molecules. *Phys. Rev. A* **66**, 051404(R) (2002).
17. Sakai, H. & Miyazaki, K. High-order harmonic-generation in nitrogen molecules with subpicosecond visible dye-laser pulses. *Appl. Phys. B* **61**, 493–498 (1995).
18. Liang, Y., Augst, S., Chin, S. L., Beaudoin, Y. & Chaker, M. High harmonic generation in atomic and diatomic molecular gases using intense picosecond laser pulses—a comparison. *J. Phys. B* **27**, 5119–5130 (1994).
19. Lyngå, C., L'Huillier, A. & Wahlström, C.-G. High-order harmonic generation in molecular gases. *J. Phys. B* **29**, 3293–3302 (1996).
20. Muth-Böhm, J., Becker, A. & Faisal, F. H. M. Suppressed molecular ionization for a class of diatomics in intense femtosecond laser fields. *Phys. Rev. Lett.* **85**, 2280–2283 (2000).
21. Tong, X. M., Zhao, Z. X. & Lin, C. D. Theory of molecular tunneling ionization. *Phys. Rev. A* **66**, 033402 (2002).
22. Itatani, J. *et al.* Tomographic imaging of molecular orbitals. *Nature* **432**, 867–871 (2004).
23. Kaku, M., Masuda, K. & Miyazaki, K. Observation of revival structure in femtosecond-laser-induced alignment of N₂ with high-order harmonic generation. *Jpn. J. Appl. Phys.* **43**, L591–L593 (2004).
24. Dooley, P. W. *et al.* Direct imaging of rotational wave-packet dynamics of diatomic molecules. *Phys. Rev. A* **68**, 023406 (2003).
25. de Nalda, R. *et al.* Role of orbital symmetry in high-order harmonic generation from aligned molecules. *Phys. Rev. A* **69**, 031804(R) (2004).
26. Young, T. Experiments and calculations relative to physical optics. *Phil. Trans. R. Soc. Lond.* **94**, 1–16 (1804).
27. Shan, B., Ghimire, S. & Chang, Z. Effect of orbital symmetry on high-order harmonic generation from molecules. *Phys. Rev. A* **69**, 021404(R) (2004).
28. Sakai, H. & Miyazaki, K. Effect of multiphoton ionization on high-order harmonic generation and propagation in rare gases with subpicosecond laser pulses. *Phys. Rev. A* **50**, 4204–4211 (1994).

Acknowledgements This work was supported by Grants-in-Aid from the Japan Society for the Promotion of Science and from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.S. (hsakai@phys.s.u-tokyo.ac.jp).

Enhanced current transport at grain boundaries in high- T_c superconductors

R. F. Klie¹, J. P. Buban², M. Varela³, A. Franceschetti^{3,4}, C. Jooss⁵, Y. Zhu¹, N. D. Browning⁶, S. T. Pantelides^{3,4} & S. J. Pennycook^{3,4}

Large-scale applications of high-transition-temperature (high- T_c) superconductors, such as their use in superconducting cables, are impeded by the fact that polycrystalline materials (the only practical option) support significantly lower current densities than single crystals^{1–6}. The superconducting critical current density (J_c) across a grain boundary drops exponentially if the misorientation angle exceeds 2° – 7° . Grain texturing reduces the average misorientation angle, but problems persist^{7,8}. Adding impurities (such as Ca in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$; YBCO) leads to increased J_c (refs 9, 10), which is generally attributed to excess holes introduced by Ca^{2+} substituting for Y^{3+} (ref. 11). However, a comprehensive physical model for the role of grain boundaries and Ca doping has remained elusive. Here we report calculations, imaging and spectroscopy at the atomic scale that demonstrate that in poly-crystalline YBCO, highly strained grain-boundary regions contain excess O vacancies, which reduce the local hole concentration. The Ca impurities indeed substitute for Y, but in grain-boundary regions under compression and tension they also replace Ba and Cu, relieving strain and suppressing O-vacancy formation. Our results demonstrate that the ionic radii are more important than their electronic valences for enhancing J_c .

The enhancement of grain-boundary J_c in poly-crystalline YBCO has now been widely reported, induced by Ca doping of both weak-linked high-angle^{9,10}, and strongly-coupled low-angle, grain boundaries^{12,13}. However, Ca doping has been shown to reduce T_c . The ideal dopant, which has yet to be found, would improve the grain-boundary J_c , but not adversely affect other superconducting properties, such as the T_c of the grains. It is usually assumed that the doping mechanism is electronic in nature, resulting in reduction of the intrinsic grain-boundary charge^{9,10,14}, modification of the bulk screening length^{9,10} or reduction of both charge and strain fields^{13,15}. However, a comprehensive atomic-scale model has not yet been found that can explain the impact of grain boundaries and Ca impurities on the critical current density. Single-crystal studies show that Ca^{2+} substitutes for Y^{3+} where it acts as a hole dopant¹¹. Song *et al.* (personal communication, and ref. 16) have found that Ca segregates in grain boundaries, and it was generally believed that Ca replaces Y there as well. However, strains of more than 10% can occur at the dislocation cores that comprise grain boundaries, with severe consequences. For example, SrTiO_3 grain boundaries are intrinsically non-stoichiometric owing to such high strains^{17,18}. Similar effects are expected in YBCO.

The formation energies required for substituting an isolated Ca dopant on different lattice sites in YBCO (in an oxidizing environment) as a function of biaxial strain in the a – b plane were calculated from first principles, and are shown in Fig. 1a. In unstrained YBCO,

Ca substitutes for Y as expected¹⁹. In regions under compressive strain, however, substitution for Ba is increasingly possible and becomes energetically favourable at strains greater than $\sim 6\%$. Similarly, in regions under tensile strain, substitution for Cu becomes favourable. In contrast, strain has only a small effect on substitution for Y. Clearly, the relative ionic sizes drive the defect formation

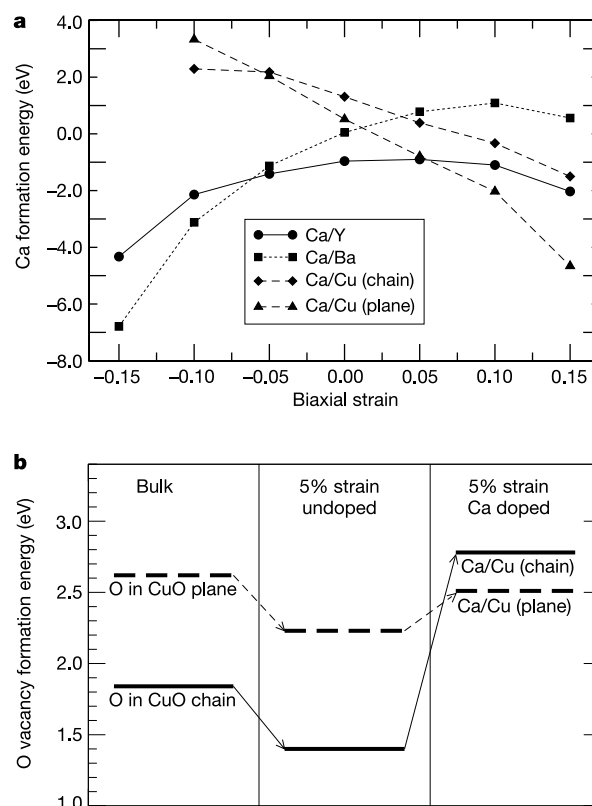


Figure 1 | First-principles calculations of Ca and O vacancy formation energy in bulk YBCO. a, Energy required to substitute Ca on different lattice sites in YBCO as a function of biaxial strain. It can be energetically favourable for Ca to replace Y, Ba or Cu, depending on the local strain. **b**, Formation energy of O vacancies in bulk YBCO (left panel), in undoped YBCO under 5% tensile strain in the a – b plane (central panel), and in Ca-doped YBCO under 5% tensile strain (right panel), simulating the expanded regions at grain boundaries. For doped grain boundaries where Ca substitutes Cu, the O vacancy formation energy increases significantly.

¹Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, USA. ²Institute of Engineering Innovation, The University of Tokyo, Tokyo, 113-8656, Japan. ³Condensed Matter Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. ⁴Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA. ⁵Institut für Materialphysik, University of Göttingen, 37073 Göttingen, Germany. ⁶Department of Chemical Engineering and Materials Science, University of California-Davis, Davis, California 95616, USA.

energies: the ionic radii of Ca and Y are roughly equal, but the radius of Cu is $\sim 30\%$ smaller and the radius of Ba is $\sim 30\%$ larger. The chemical valence of impurities becomes secondary in the presence of these high strains. A potentially unique choice as an alternative dopant could therefore be Ag, which has roughly the same ionic radius as Y and Ca, and can also be isovalent with Y so that it would not have any direct effect on the hole concentration in the grains. These dopants might increase the superconducting current across the grain boundary like Ca, but without detrimental effects on superconducting properties inside the grain.

We conclude that Ca substitution at the grain boundary occurs in order to reduce the local strain. In turn, the concomitant strain relief has a major effect on the O vacancy formation energies in the vicinity. Our calculations show that in bulk YBCO, O vacancies form preferentially along the CuO chains, with a formation energy of ~ 1.9 eV (Fig. 1b); the O vacancy formation energy in the CuO₂ planes is ~ 0.7 eV higher. Moreover, Fig. 1b shows that in strained, undoped YBCO, the formation energies for O vacancies are significantly reduced, both in the CuO₂ planes and CuO chains. O vacancies would therefore segregate to an undoped grain boundary, explaining the oxygen deficiency and hole depletion known from previous studies^{20,21}. When Ca doping is introduced, O vacancy formation

energies increase, becoming even higher than in the bulk in the case of the CuO chains. Hence, compared to the undoped grain boundary, the doped grain boundary should show much reduced deficiency of O, and therefore reduced hole depletion. Thus, our results confirm the fact that excess O vacancies in grain boundaries and their corollary effects are the primary reason for the observed reduction in critical currents in poly-crystalline YBCO. The role of Ca doping is to remove the intrinsic excess of O vacancies from YBCO grain boundaries. Thus, high critical currents are restored through a cooperative doping mechanism.

To test the theoretical predictions experimentally, we measured electron energy-loss spectra (EELS) in an aberration-corrected scanning transmission electron microscope (STEM). With a probe size around 1 Å, such instruments are able to provide 'column-by-column' spectroscopy, and directly confirm the effects predicted by theory. Thin films of YBCO, both pristine and 20% Ca doped, were grown by laser ablation on SrTiO₃ bicrystal substrates at the University of Göttingen. Superconducting properties were analysed previously, and a 35% increase in J_c was found on Ca doping¹³.

The dislocation core structures were found to be quite different in the Ca-doped case as compared to the undoped sample. As shown in Fig. 2, the Ca-doped core shows a complete additional atomic

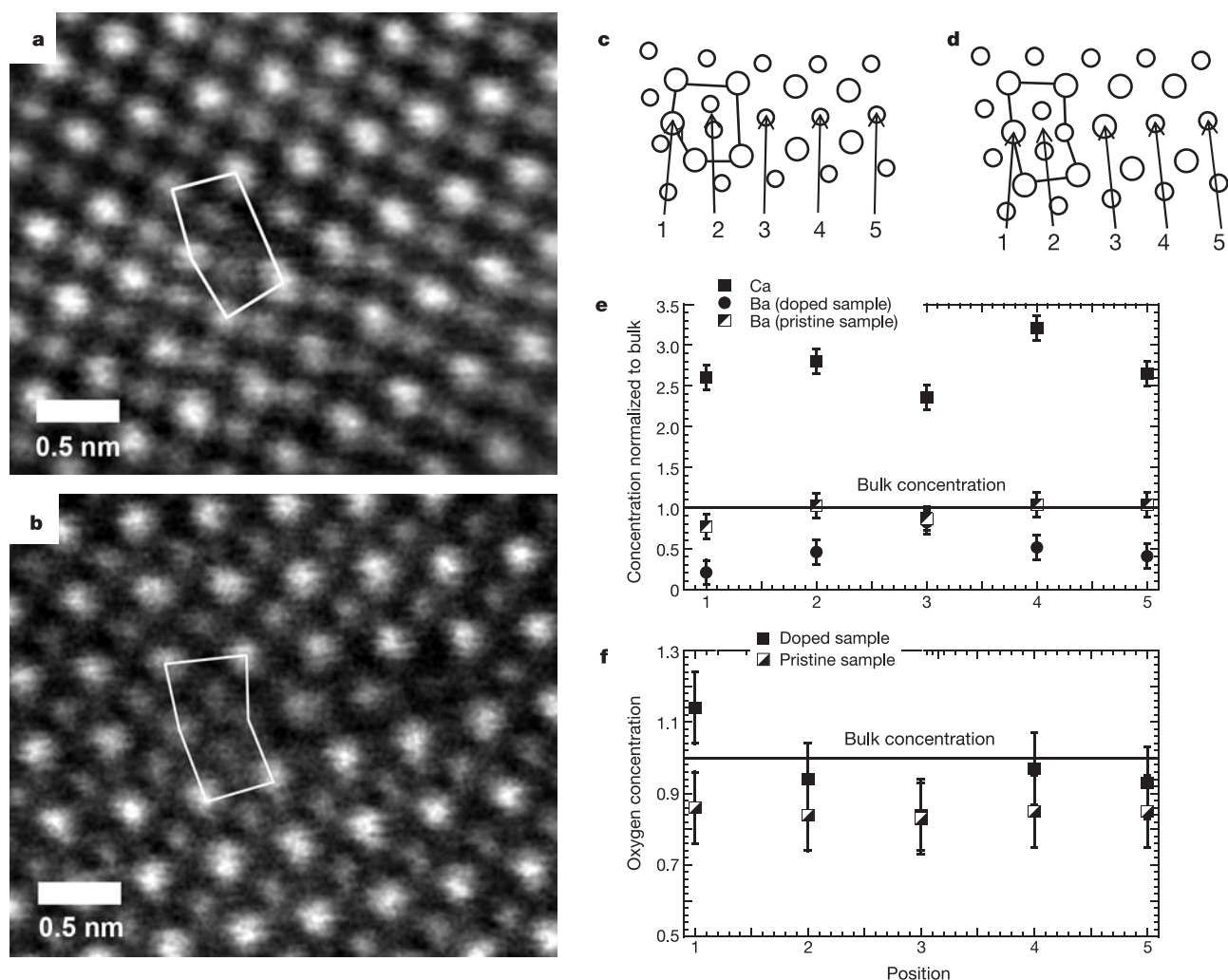


Figure 2 | Structural differences in pristine and Ca-doped YBCO. **a**, Atomic-resolution Z-contrast image of a pristine 4° [001] tilt grain-boundary dislocation core on the Cu-O sublattice, showing the typical pentagonal arrangement²². **b**, A Ca-doped grain boundary: the Y/Ba column pentagonal arrangement encloses three columns, two on the Cu-O and one on the Y/Ba

sublattice. **c** and **d**, schematic core structure of **a** and **b**, respectively. **e**, Profile of Ca and Ba concentration for probe positions 1–5 indicated in **c** and **d**, normalized to the bulk. **f**, O K-edge quantification for probe positions 1–5, showing less O-deficiency in the doped dislocation cores (error bars indicate the standard error on the concentration quantification).

column at its centre. The column pairs in the undoped dislocation cores are typically seen in perovskite grain boundaries, but in the doped boundary there are instead three columns. These structures were seen consistently over the entire grain boundary. In addition, the lattice perpendicular to the grain boundary is significantly expanded in the Ca-doped sample (from 17% in the pristine core to 53% in the doped core). This expanded region extends for several unit cells, to the right of the core itself. EELS data were taken from the probe positions numbered in the schematic (Fig. 2c and d). The probe size was unchanged from that used for imaging, and therefore provides a column-by-column analysis with the same resolution as the image^{22–24}. In the Ca-doped sample, the average increase in Ca concentration was ~ 2.5 times compared to the bulk, as shown in Fig. 2e. On the right of the core, the columns are nominally Cu–O columns, consistent with the theoretical prediction that Ca should replace Cu in the expanded lattice. At the left of the core, column 1 is nominally a Y/Ba column, and the presence of Ca in this compressed region is again consistent with the theoretical prediction. Furthermore, of the five columns analysed, only column 3 showed a statistically significant decrease in O concentration, as shown in Fig. 2f. All other columns showed an increase in O concentration compared to the undoped dislocation core, which showed 15–20% depletion in O from each of the columns 1–5.

Besides the determination of elemental concentrations, EELS provides a quantitative measure of the carrier concentration from the fine structure on the O K-edge, specifically from the ratio of the pre-peak to the main peak^{20,21}. The O pre-peak intensity (see Supplementary Figs S1 and S3) showed substantial reduction in the undoped dislocation core, confirming the expected carrier depletion. In the Ca-doped core (see Supplementary Fig. S2), however, the reduction was much smaller, although still large compared to the bulk value, indicating that significant hole-depletion remains.

The EELS data also explain several consistent changes to the image contrast in the doped core. In all cores observed, column 1, nominally an Y/Ba column, was darker than equivalent columns in the bulk. EELS shows this column to have the highest Ca/Ba ratio, double the bulk value. Because the image contrast is approximately proportional to the square of the atomic number (Z), replacing Ca ($Z = 20$) for Ba ($Z = 56$) would reduce the brightness of the column. On the other hand, column 3, nominally a CuO column, was particularly bright. This column showed the highest Ba concentration (Fig. 2e). Ba is isovalent with Ca, so it is consistent with the principle of maximizing strain relief to replace Ba, with the longest bond to O, in the most highly strained unit cell. Calculations confirmed that Ba substituting for Cu in the chain is also energetically favourable, and again raises the O vacancy formation energy to near its bulk value.

Finally, we compare the present physical model of the role of grain boundaries and Ca doping in YBCO with earlier work. Our model is consistent with the conventional view that grain boundaries in YBCO contain O vacancies, which act as donors and result in positive charge in the grain-boundary plane (the energy bands bend down). The positive charge sets up a space-charge region of hole depletion caused by excess screening electrons. We have shown that, through the cooperative doping mechanism, both the grain-boundary charge and the space-charge region are reduced. These results contrast only with a recent theory paper by Su and Welch¹⁹ and the holography data from Schofield *et al.*¹⁴. Both these papers concluded that undoped grain boundaries are negatively charged (bands bend up and are completely empty near the grain boundaries, that is, mobile carriers are again depleted) and doping mitigates the effect. We can trace the results of Su and Welch to the fact that they performed calculations for a free surface, and assumed that grain boundaries behave the same way. In addition, these authors only considered Ca substituting for Y. In the case of the holography data, the reported sign of the grain-boundary potential is negative. The discrepancy raises

questions about what potential is measured by the holography beam, but these issues are beyond the scope of this Letter. We only note that the question is not settled, as an earlier paper on dislocations in GaN reported a negative potential that could not be accounted for²⁵.

METHODS

Atomic resolution Z-contrast imaging and EELS. We use atomic resolution Z-contrast imaging in the STEM to show the arrangement of the cation columns at the boundary, and correlated EELS to probe the concentration of impurities such as Ca and the local concentration of holes^{20–24}. The Z-contrast images were collected using the aberration corrected VG-HB603U with an optimum probe size of 0.6 Å (ref. 26) and a detector inner angle of ~ 40 mrad. In these collection conditions, the image contrast of an atomic column in the incoherent Z-contrast image is approximately proportional to the square of the average atomic number (Z). This means that the bright spots in Fig. 2 represent the Y/Ba columns, while the less bright columns are Cu–O of the YBCO [001] projection; the O columns are not visible in these micrographs. The additional spot in the Ca-doped dislocation core centre (Fig. 2b) can be directly interpreted as an additional atomic column on the Y/Ba/Ca sublattice, thereby transforming the pentagonal dislocation core in the Ca-doped grain boundary into two bulk-like unit cells under compressive strain.

The correlated energy-loss spectra were acquired using the VG-HB501UX, an aberration-corrected dedicated STEM instrument with a cold-field emission electron gun, operated at 100 kV, and equipped with a Gatan Enfina spectrometer, using an energy dispersion of 0.3 eV per channel, thus yielding a nominal spatial resolution of 1.2 Å and a spectral resolution of 0.4 eV (ref. 24). The spectra were acquired from the bulk and at a specific position in the grain boundary dislocation core structure to measure changes in the Ca L-edge, O K-edge and the Ba M-edge as a function of position in the grain-boundary dislocations. As the excitation energies for the Ca L-edge, the O K-edge and the Ba M-edge are near each other, all three edges can be recorded in a single spectrum. For each position, spectra were acquired for 5 s and several spectra were acquired from similar positions in different dislocation cores to obtain a good signal-to-noise ratio. Sample drift and damage were also checked after each spectrum acquisition. The spectra were then summed and subsequently background-subtracted²⁷.

Density-functional calculations. Calculations of O vacancy and Ca impurity formation energies were performed using density-functional theory in the local-density approximation²⁸. First-principles calculations for realistic, low-angle YBCO grain-boundary structures are computationally prohibitive. As the driving force for defect segregation at the grain boundary is the local strain field, we performed calculations of defect formation energies in biaxially strained YBCO. Ultra-soft pseudopotentials were used to describe electron–ion interactions, and the electronic wavefunctions were expanded in a plane-wave basis set with an energy cut-off of 29 Rydberg. A periodically repeated 52-atom supercell (corresponding to 4 bulk unit cells) was used to minimize defect–defect interactions. All atomic positions were fully relaxed until the quantum-mechanical forces acting on the atoms were negligible.

The formation energies of Ca substitutional impurities were calculated for oxygen-rich conditions, using different amounts of compressive and tensile biaxial lattice strain to mimic the local strain in different regions of the grain boundaries. The formation energy of a Ca impurity substituting for a metal atom A ($A = \text{Y, Cu or Ba}$) in YBCO is given by $\Delta F_{\text{Ca}/A} = -\mu_{\text{Ca}} + \Delta E_{\text{Ca}/A} + \mu_{\text{A}}$, where $\Delta E_{\text{Ca}/A} = E_{\text{Ca}} - E_{\text{A}}$ is the energy required to replace an A atom with a Ca atom in the YBCO supercell, and μ_{Ca} and μ_{A} are the chemical potentials of Ca and A. In oxygen-rich conditions, μ_{A} (and μ_{Ca}) are given by $\mu_{\text{A}} = [E(\text{A}_m\text{O}_n) - n/2 E(\text{O}_2)]/m$, where $E(\text{A}_m\text{O}_n)$ is the formation energy of the oxide A_mO_n (the most stable oxide in the presence of molecular oxygen) and $E(\text{O}_2)$ is the binding energy of the O_2 molecule.

Received 2 December 2004; accepted 14 April 2005.

- Dimos, D., Chaudhari, P. & Mannhart, J. Superconducting transport properties of grain boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_7$ bicrystals. *Phys. Rev. B* **41**, 4038–4049 (1990).
- Gross, R. & Mayer, B. Transport processes and noise in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ grain-boundary junctions. *Physica C* **180**, 235–242 (1991).
- Gurevich, A. & Pashitskii, E. A. Current transport through low-angle grain boundaries in high-temperature superconductors. *Phys. Rev. B* **57**, 13878–13893 (1998).
- Heinig, N. F. *et al.* Evidence for channel conduction in low misorientation angle [001] tilt $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ bicrystal films. *Appl. Phys. Lett.* **69**, 577–579 (1996).
- Hilgenkamp, H. & Mannhart, J. Grain boundaries in high- T_c superconductors. *Rev. Mod. Phys.* **74**, 485–549 (2002).

6. Mannhart, J., Chaudhari, P., Dimos, D., Tsuei, C. C. & McGuire, T. R. Critical currents in [001] grains and across their tilt boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_7$ films. *Phys. Rev. Lett.* **61**, 2476–2479 (1988).
7. Iijima, Y. *et al.* Structural and transport-properties of biaxially aligned $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ films on polycrystalline Ni-based alloy with ion-beam-modified buffer layers. *J. Appl. Phys.* **74**, 1905–1911 (1993).
8. Norton, D. P. *et al.* Epitaxial $\text{YBa}_2\text{Cu}_3\text{O}_7$ on biaxially textured nickel (001): An approach to superconducting tapes with high critical current density. *Science* **274**, 755–757 (1996).
9. Hammerl, G. *et al.* Enhanced supercurrent density in polycrystalline $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ at 77 K from calcium doping of grain boundaries. *Nature* **407**, 162–164 (2000).
10. Schmehl, A. *et al.* Doping-induced enhancement of the critical currents of grain boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. *Europhys. Lett.* **47**, 110–115 (1999).
11. Kucera, J. T. & Bravman, J. C. Transport characterization of calcium-doped $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ thin films. *Phys. Rev. B* **51**, 8582–8590 (1995).
12. Daniels, G. A., Gurevich, A. & Larbalestier, D. C. Improved strong magnetic field performance of low angle grain boundaries of calcium and oxygen overdoped $\text{YBa}_2\text{Cu}_3\text{O}_x$. *Appl. Phys. Lett.* **77**, 3251–3253 (2000).
13. Guth, K., Krebs, H. U., Freyhardt, H. C. & Jooss, C. Modification of transport properties in low-angle grain boundaries via calcium doping of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ thin films. *Phys. Rev. B* **64**, 140508 (2001).
14. Schofield, M. A., Beleggia, M., Zhu, Y. M., Guth, K. & Jooss, C. Direct evidence for negative grain boundary potential in Ca-doped and undoped $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. *Phys. Rev. Lett.* **92**, 195502 (2004).
15. Kung, H. *et al.* Dissociation of grain boundary dislocations in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ -coated conductors. *Phil. Mag. Lett.* **81**, 85–93 (2001).
16. Song, X., Daniels, G. A., Feldmann, D. M., Gurevich, A. & Larbalestier, D. C. Electromagnetic, atomic-structure and chemistry changes induced by Ca-doping of low angle YBCO grain boundaries. *Nature Mater.* (submitted).
17. Kim, M. *et al.* Nonstoichiometry and the electrical activity of grain boundaries in SrTiO_3 . *Phys. Rev. Lett.* **86**, 4056–4059 (2001).
18. Klie, R. F., Beleggia, M., Zhu, Y., Buban, J. P. & Browning, N. D. Atomic-scale model of the grain boundary potential in perovskite oxides. *Phys. Rev. B* **68**, 214101 (2003).
19. Su, H. B. & Welch, D. O. Effects of space charge, dopants, and strain fields on surfaces and grain boundaries in YBCO compounds. *Supercond. Sci. Technol.* **18**, 24–34 (2005).
20. Browning, N. D. *et al.* The atomic origins of reduced critical currents at [001] tilt grain boundaries in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ thin films. *Physica C* **294**, 183–193 (1998).
21. Zhu, Y., Zuo, J. M., Moodenbaugh, A. R. & Suenaga, M. Grain boundary constraint and oxygen deficiency in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. *Phil. Mag. A* **70**, 969–984 (1994).
22. Allen, L. J., Findlay, S. D., Lupini, A. R., Oxley, M. P. & Pennycook, S. J. Atomic-resolution electron energy loss spectroscopy imaging in aberration corrected scanning transmission electron microscopy. *Phys. Rev. Lett.* **91**, 105503 (2003).
23. Browning, N. D., Chisholm, M. F. & Pennycook, S. J. Atomic-resolution chemical analysis using a scanning transmission electron microscope. *Nature* **366**, 143–146 (1993).
24. Varela, M. *et al.* Spectroscopic imaging of single atoms within a bulk solid. *Phys. Rev. Lett.* **92**, 095502 (2004).
25. Cherns, D. & Jiao, C. G. Electron holography studies of the charge on dislocations in GaN. *Phys. Rev. Lett.* **87**, 205504 (2001).
26. Nellist, P. D. *et al.* Direct sub-angstrom imaging of a crystal lattice. *Science* **305**, 1741 (2004).
27. Egerton, R. F. *Electron Energy Loss Spectroscopy in the Electron Microscope* (Plenum, New York, 1986).
28. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D.W. Gillette for the sample preparation and A.R. Lupini for his help in acquiring the images of Ca-doped YBCO. This work is supported by the US Department of Energy, Division of Materials Sciences, Office of Basic Energy Science and the National Energy Research Scientific Computing Center, supported by the Office of Science of the US Department of Energy.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.F.K. (klie@bnl.gov).

Changes in carbon dioxide during an oceanic anoxic event linked to intrusion into Gondwana coals

Jennifer C. McElwain¹, Jessica Wade-Murphy² & Stephen P. Hesselbo³

The marine sedimentary record exhibits evidence for episodes of enhanced organic carbon burial known as 'oceanic anoxic events' (OAEs)^{1,2}. They are characterized by carbon-isotope excursions in marine³ and terrestrial⁴ reservoirs and mass extinction of marine faunas⁵. Causal mechanisms for the enhancement of organic carbon burial during OAEs are still debated^{6,7}, but it is thought that such events should draw down significant quantities of atmospheric carbon dioxide⁷. In the case of the Toarcian OAE (~183 million years ago), a short-lived negative carbon-isotope excursion in oceanic and terrestrial reservoirs has been interpreted to indicate raised atmospheric carbon dioxide⁴ caused by oxidation of methane catastrophically released from either marine gas hydrates⁴ or magma-intruded organic-rich rocks⁸. Here we test these two leading hypotheses^{4,8} for a negative carbon isotopic excursion marking the initiation of the Toarcian OAE using a high-resolution atmospheric carbon dioxide record obtained from fossil leaf stomatal frequency^{9,10}. We find that coincident with the negative carbon-isotope excursion carbon dioxide is first drawn down by 350 ± 100 p.p.m.v. and then abruptly elevated by $1,200 \pm 400$ p.p.m.v. and infer a global cooling and greenhouse warming of 2.5 ± 0.1 °C and 6.5 ± 1 °C, respectively. The pattern and magnitude of carbon dioxide change are difficult to reconcile with catastrophic input of isotopically light methane from hydrates⁵ as the cause of the negative isotopic signal. Our carbon dioxide record better supports a magma-intrusion hypothesis⁸, and suggests that injection of isotopically light carbon from the release of thermogenic methane occurred owing to the intrusion of Gondwana coals by Toarcian-aged Karoo-Ferrar dolerites.

The Toarcian OAE is marked by worldwide deposition of organic-rich black shale, and represents a major perturbation in the global carbon cycle. We have reconstructed a high-resolution and stratigraphically well-constrained record of the partial pressure of atmospheric CO₂, p_{CO_2} , spanning the Toarcian OAE. We estimated p_{CO_2} using the stomatal proxy method^{9,10} on 126 fossil leaves collected as dispersed mesofossils, preserved within nearshore sediments of the Sorthat (formerly Bagå) Formation at Korsodde, Bornholm, in the eastern Danish basin^{11,12}. This approach to p_{CO_2} reconstruction uses a genetically controlled inverse relationship between p_{CO_2} and stomatal index (SI; see Methods). The method has a demonstrated accuracy of ± 20 p.p.m.v. and ± 30 p.p.m.v. in the Quaternary¹³ and Neogene¹⁰, respectively, and is unaffected by changes in temperature, precipitation, leaf size, or any environmental or biological factors other than p_{CO_2} or (to a lesser extent) light intensity¹⁴.

SI data were obtained following standard protocols¹³ from charcoalified and coalified mesofossil leaves of Coniferales (*Brachyphyllum*, *Pagiophyllum*, *Cyparissidium*), Ginkgoales (*Baiera*, *Ginkgoites*), Czekanowskiales (*Czekanowskia*, *Solenites*, *Hartzia*), Pteridosperms (*Pachypteris*, *Sagenopteris*) and Cycadales, collected at 20 cm

stratigraphic sample resolution (of the order of 10 kyr) from strata equivalent to the upper *tenuicostatum* and lower *falciferum* ammonite zones of the Toarcian (~183 Myr ago¹⁵). The Korsodde section (Fig. 1) has previously been shown to exhibit a clear isotopic anomaly in the contained terrestrial organic matter⁴. We have divided the anomaly into two parts: an initial interval of gradual decline in carbon-isotope values (phase A), which coincides with the initiation of the Toarcian OAE, and a later phase characterized by a shift to even lighter isotopic values (phase B) and peak burial of organic carbon in stratigraphically correlated marine sections⁴. The negative isotopic anomaly from carbonate and organic matter occurs in widely distributed sections and is thought to be a global phenomenon³.

Mean leaf SIs from 126 mesofossil leaves were statistically smoothed using a locally weighted polynomial regression (LOESS¹⁶) to determine an ecosystem-level (coastal and local upland floras) response to changes in p_{CO_2} (Fig. 1, Supplementary Table 1). The resultant smoothed SI records indicate (1) a pattern of gradually declining SI preceding the OAE, (2) a marked twofold increase in SI during phase A, and (3) a 60% decrease in SI within the OAE (across phase A to B). Random resampling (Supplementary Table 2) demonstrated that the smoothed SI trends are not dictated by sample density (which varies widely throughout the section), confirming their robustness (Fig. 1).

Toarcian p_{CO_2} changes (Fig. 2) were quantified from ecosystem level SI records (Fig. 1, Supplementary Table 1) by extracting two subsets of SI data from fossil Ginkgoales and Coniferales, as both families are known to be highly CO₂ sensitive^{9,13,17} and both are represented by suitable nearest living equivalent (NLE) taxa necessary to calibrate p_{CO_2} from fossil stomatal responses. The resultant p_{CO_2} records (see Methods) obtained using different calibration approaches from two independent families reveal remarkably congruent fluctuations through the Toarcian (Fig. 2b). Calibrations based on Coniferales indicate that p_{CO_2} , and by inference global mean land surface temperature (GMLST) expressed as ΔT (temperature departure from present day; see Methods), gradually increased in the *tenuicostatum* zone before the OAE from 600 ± 150 p.p.m.v. to $\sim 950 \pm 230$ p.p.m.v. and by about +2 °C, respectively. This trend is supported by independent p_{CO_2} estimates from Ginkgoales SIs, which demonstrate a p_{CO_2} rise from 475 ± 125 p.p.m.v. to $1,040 \pm 260$ p.p.m.v. (Fig. 2b) over the same interval (26–31 m). Following a gap in the record (of the order of 300 kyr) between 30.9 and 35.2 m (due to an absence of well-preserved fossil leaves), p_{CO_2} and GMLST show a marked decline of $\sim 350 \pm 105$ p.p.m.v. and $\sim 2.5 \pm 0.1$ °C (Fig. 2) respectively in phase A, coincident with the local (and probably global) onset of organic-rich shale deposition¹⁸ marking the Toarcian OAE. During this interval, we estimate that p_{CO_2} may have been as low as 360 p.p.m.v. and in the general range of 400–600 p.p.m.v. for an

¹Department of Geology, The Field Museum, 1400 S. Lake Shore Drive, Chicago, Illinois 60605-2496, USA. ²Department of Ecology and Evolution, University of Chicago, 5801 South Ellis Avenue, Chicago, Illinois 60637, USA. ³Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

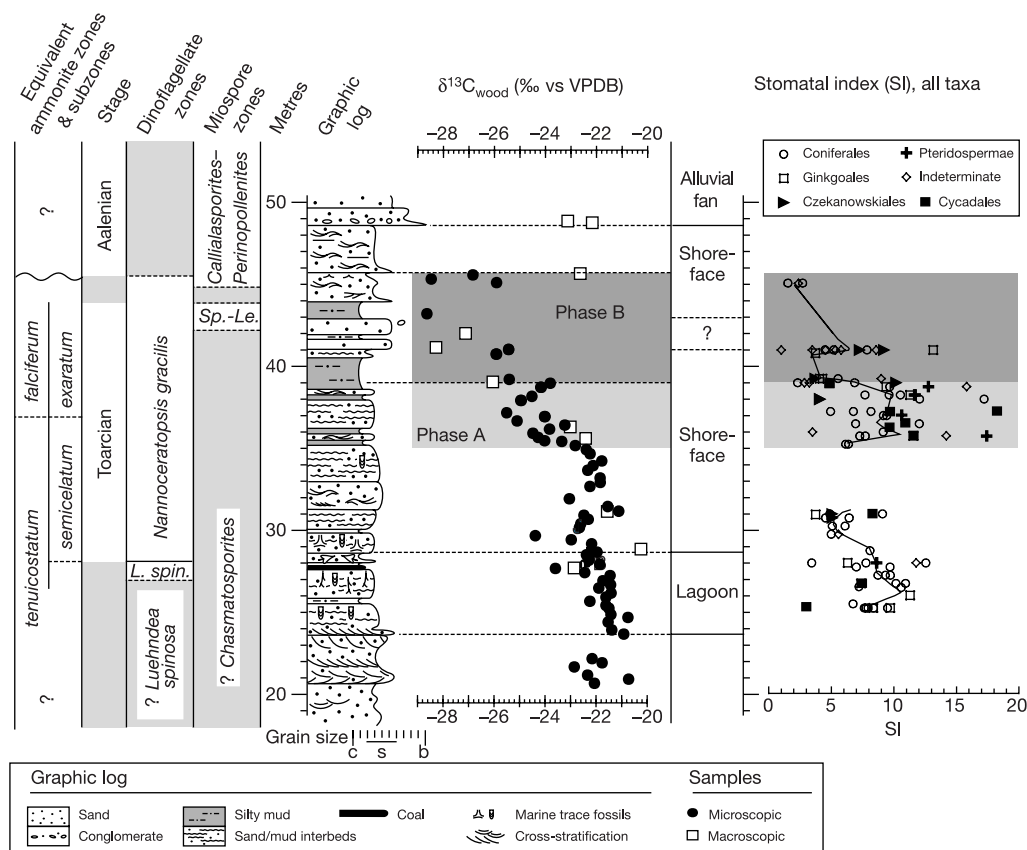


Figure 1 | Carbon isotope and stomatal index data through the Sorthat Formation. The Sorthat Formation¹² (formerly Bagå Formation¹¹), is located in southwest Bornholm, Denmark. Carbon-isotope data, ammonite stratigraphy and stage assignments are summarized from ref. 4. Stomatal index (SI) data are from charcoalified or coalified leaf mesofossils of Coniferales, Ginkgoales, Pteridospermales and Czekanowskiales. Data points represent mean ($n = 3$ to 5) SI values per mesofossil leaf specimen,

and the line represents best fit using a locally weighted polynomial regression (LOESS¹⁶). Depositional environments are summarized from ref. 11, although the beds from 40–43 m are of uncertain origin (exposure is limited and diagnostic features lacking). 'Phase A' and 'phase B' refer to initial and peak phases of the negative C-isotope excursion, and are represented by pale and dark shades of grey in this figure and in Fig. 2. Grain size scale: c, clay; s, sand; b, boulder.

estimated 200,000 yr (see Methods). This apparently large p_{CO_2} drawdown and global cooling was terminated by an abrupt $\sim 1,200 \pm 400$ p.p.m.v. p_{CO_2} increase and a $\sim 6.5 \pm 1^\circ\text{C}$ inferred greenhouse warming which occurred within 1 m (of the order of 50 kyr) across the transition from phase A to B.

Independent mid-latitude sea surface temperature estimates based on Belemnite Mg/Ca ratios and $\delta^{18}\text{O}$ indicate a cool *tenuicostatum* zone followed by a warm *falciferum* zone; the temperature change between extremes being about $15\text{--}20^\circ\text{C}$ (refs 3, 19), in agreement with our GMLST record. Mesofossil and pollen relative abundances within the Sorthat Formation indicate ecological dominance by the thermophilic conifer *Pagiophyllum* and its related pollen *Corollina torosus*¹¹ (a known high temperature/aridity indicator²⁰) before the OAE, with declines in phase A (Fig. 2d) providing independent support for a CO_2 mediated transient cooling during the initial stages of the OAE. The p_{CO_2} spike is coeval with a large positive $^{187}\text{Os}/^{188}\text{Os}$ excursion in the Jet Rock of Yorkshire, hypothesized to represent a transient increase in global continental weathering rates by an estimated 400–800% (ref. 21), an expected feedback response to greenhouse induced warming. Following this spike, atmospheric p_{CO_2} and global temperatures returned to pre-excursion levels. Importantly, the identified trends in p_{CO_2} do not correlate with any facies change within the Sorthat Formation, and are detected within two independent families (Coniferales and Ginkgoales), ruling out the likelihood that they are driven by any undetected species level changes in the composition of vegetation.

Neither the magnitude nor the pattern of p_{CO_2} change identified from our stomatal analysis are easy to reconcile with the methane hydrate hypothesis⁴ as the explanation for a negative carbon isotopic anomaly ($> -2\text{‰}$) during the Toarcian OAE. We estimate that 2,600–4,400 Gt of carbon were rapidly incorporated into the atmosphere as CO_2 at the start of phase B (see Methods). These estimates are much larger ($> 10,000$ Gt) if coeval uptake by the oceans is taken into account²². On the basis of recent estimates that significantly downgrade the size of the ancient gas hydrate reservoir from 15,000 Gt to between 500 and 2,500 Gt (ref. 23), it is unlikely that such a large injection of isotopically light carbon could be solely derived from a methane hydrate source. Further, the detected $\sim 200,000$ yr episode of surprisingly low p_{CO_2} , and by inference globally cooler climates during the initial part of the C isotopic excursion (phase A), is not compatible with a volcanic CO_2 induced global warming mechanism (resulting in increased bottom water temperatures) as the leading hypothesis for methane hydrate dissociation⁴.

Although multiple forcing factors may have contributed to the abrupt spike in atmospheric CO_2 , we suggest that the likely dominant forcing factor was oxidation of methane gas generated by subsurface thermal metamorphism of organic-rich late Permian and late Triassic coal bearing strata^{24,25} during magmatic intrusion of the Karoo-Ferrar large igneous province (LIP) of southern Gondwana⁸. These coals are a characteristic and globally significant deposit in high southern latitude Gondwana²⁴. Peak volcanic activity of this LIP is

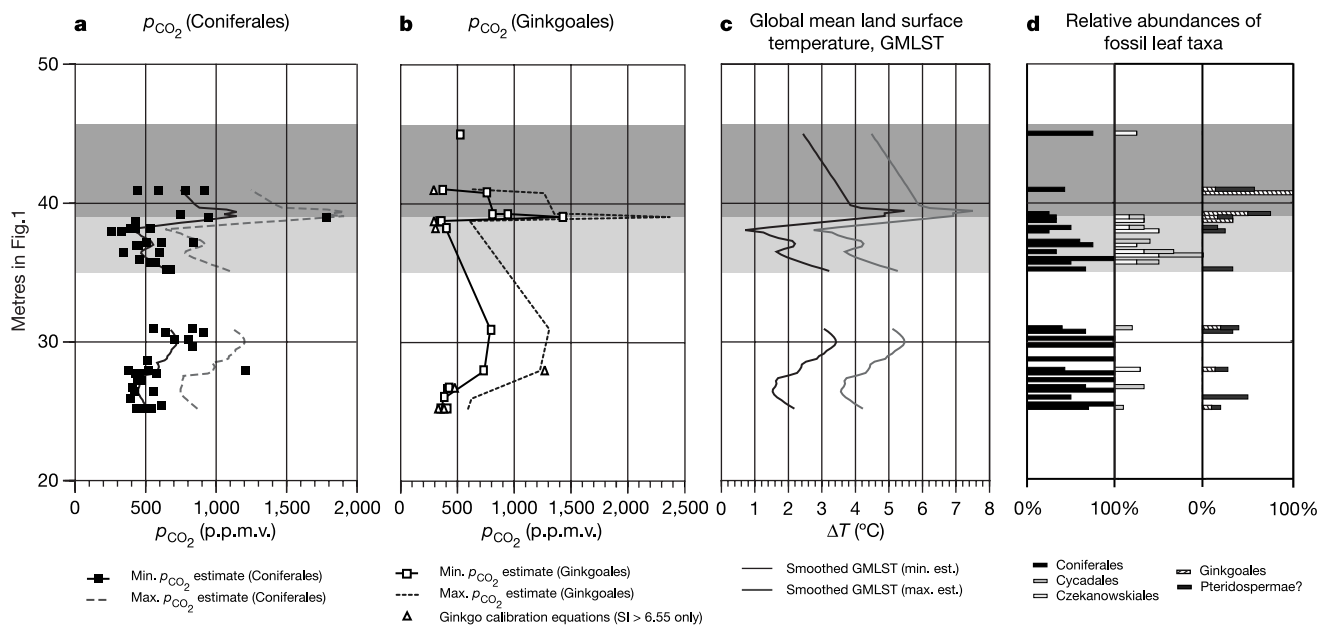


Figure 2 | Carbon dioxide, global temperature and palaeoecological trends through the Toarcian OAE. **a, b,** Reconstructed atmospheric CO_2 records (p_{CO_2}) for the Early Toarcian based on SI of mesofossil Coniferales leaves (**a**) and on SD of Ginkgoales (**b**) from the Sorthat Formation¹². Continuous and broken lines represent LOESS best fit¹⁶ p_{CO_2} estimates calibrated using a 'Recent standardization' and 'Carboniferous standardization' respectively⁹ (see Methods). Triangles represent independent p_{CO_2} estimates calibrated

using a *Ginkgo biloba* p_{CO_2} versus SI calibration curve¹⁷. Shaded grey boxes indicate phase A and B of the isotopic excursion as defined for Fig. 1. **c,** Model-determined temperature departure (ΔT) of global mean land surface temperature (GMLST) from the present day, calculated from p_{CO_2} estimates using a CO_2 –temperature sensitivity study³⁰. **d,** Relative abundances of leaf mesofossils sorted from bulk sediment samples through the Sorthat Formation.

dated at $\sim 183 \pm 2$ Myr ago (ref. 15), indistinguishable in age from the OAE. Dolerite sills dominate the intrusion complex that underlies the Karoo-Ferrar lavas²⁵. Intrusive magma volumes and organic carbon masses are difficult to quantify on the basis of available data. Nevertheless, where best developed in Antarctica, sills are on average 100–200 m thick and locally exceed 700 m, with cumulative thicknesses up to 1 km; individual sills have an estimated aerial extent of as much as 60,000 km² (see summary in ref. 25). Thus the magnitude of intrusion and the potential for liberation of methane from Gondwana coals ($\delta^{13}\text{C} = -35$ to -50‰ ; ref. 8) would appear to exceed that calculated for the North Atlantic LIP (which resulted in an estimated maximum release of 3,000 Gt of methane) during the Palaeocene–Eocene thermal maximum⁸. Independent evidence for large-scale generation of methane and carbon dioxide from Gondwana coals comes from the occurrence of late (replacement) cements in the Beacon Sandstone of Antarctica, which were formed from Ca ions, probably using CO_2 and CH_4 to raise the pH (ref. 26).

Global wildfire²⁷ or overturn of a previously stagnant Tethys Ocean²⁸ are very unlikely potential causal mechanisms for the p_{CO_2} spike and negative C isotopic anomaly. The calculated magnitude of our p_{CO_2} spike (from a minimum of 750 p.p.m.v. to a maximum of 1,600 p.p.m.v.) is too large to be explained by burning of the entire accumulated Toarcian terrestrial organic carbon reservoir (estimated to have been ~ 900 – $1,100$ Gt peat/soil carbon at 500 p.p.m.v. palaeoatmospheric CO_2 ; ref. 29). More importantly, charcoal abundance (indicative of palaeowildfire) decreases in the Sorthat Formation⁴ and palaeoclimatic indicators are suggestive of a wetter climate (Fig. 2d) coeval with the p_{CO_2} spike: findings that are incompatible with a global wildfire scenario. Similarly, overturn of the Tethys as a source of isotopically light C requires the development of anoxic or strongly dysaerobic facies in the deep sea record before the OAE, a phenomenon not known from the stratigraphic record¹.

Our p_{CO_2} record, in conjunction with the wood isotopic data⁴, elucidate important controls on the carbon cycle during the Toarcian

OAE. The negative excursion shows that light carbon (probably methane released from subsurface thermal metamorphism of Gondwana coals^{24,25}) was being added to the atmospheric and shallow ocean carbon reservoirs through phases A and B. Low CO_2 in phase A suggests that this input was more than compensated by removal via enhanced organic carbon burial^{1,3,4}, which, judging by the Jet Rock and other records^{1,3,4}, started up at this time. High CO_2 during the transition from phase A to B coincides with further organic carbon burial, indicated by a trend towards peak total organic carbon values in local and global records³ and a sustained osmium isotope anomaly in phase B (ref. 21), showing that although carbon sink processes were still accelerating, input of light carbon into the system via Karoo-Ferrar magmatism was also accelerating, and overwhelmed the processes for removal.

The evidence presented here supports the prevailing view of the Toarcian OAE as a time of increased warmth, but also demonstrates that enhanced organic carbon burial/preservation at the onset of the OAE resulted in atmospheric conditions ($p_{\text{CO}_2} < 400$ p.p.m.v.) that are associated with cooling, possibly of a sufficient magnitude even to enable transient ice cap growth (R. DeConto, personal communication) at higher elevations of northern and southern high latitudes of Pangaea.

METHODS

Estimating p_{CO_2} from fossil stomata. The stomatal index (SI) is defined as the ratio of the number of stomata to the total number of epidermal cells plus stomata within a given leaf area expressed as a percentage. Minimum and maximum CO_2 concentrations were first estimated following standard calibration approaches⁹, from the SI of fossil Coniferales. Coniferales were particularly well represented throughout the entire record, comprising 60 of the 126 fossil leaves analysed and absent from only 5 out of a total of 24 sampling horizons. *Pagiophyllum* was the dominant form-genus with rare occurrences of *Brachyphyllum* and one *Cyparissidium*. 'Recent' and 'Carboniferous standardizations'⁹ were applied to Coniferales stomatal ratios (calculated as the ratio of the SI of their NLE *Araucaria heterophylla* (SI = 11.4%) to fossil *Pagiophyllum* SI). Additional independent CO_2 estimates were obtained from fossil Ginkgoales

(*Baiera* and *Ginkgoites*) stomatal density (SD; number of stomata per mm² leaf area, Supplementary Table 1), using two different but well tested calibration approaches. The first used the same approach as for the Coniferales with *Ginkgo biloba* (SD = 84 mm⁻²), the only extant taxon within Ginkgoales, selected as NLE from which stomatal ratios were calculated and calibrated according to ref. 9. To test the robustness of the stomatal ratio approach to p_{CO_2} reconstruction, p_{CO_2} estimates from ginkgoalean SI were also obtained using an independent training set of *Ginkgo biloba* SI responses to both historical and experimentally altered CO₂ concentrations¹⁷. It must be noted that only fossil ginkgoalean SIs in the range of the extant *Ginkgo biloba* SI training set (SI > 6.55 and < 13.98)¹⁷ were calibrated for p_{CO_2} .

Estimating time represented by Sorthat Formation sediments. The temporal resolution of mesofossil leaf samples was roughly estimated using a timescale from ref. 15 and assuming constant sedimentation rate and complete representation of time equivalent to the upper *tenuicostatum* and lower *falciferum* zones in the Sorthat Fm⁴.

Estimating GMLST. Global mean land surface temperature (GMLST) was calculated by assuming a greenhouse relationship between p_{CO_2} and global temperatures using a CO₂-climate sensitivity study³⁰ suggesting $\Delta T = 4 \ln(R_{\text{CO}_2})$, where R_{CO_2} is the ratio of the estimated palaeo-CO₂ concentration to that of pre-industrial times (300 p.p.m.v.).

Estimating amount of carbon released during p_{CO_2} spike. Following ref. 22 (1 p.p.m.v. CO₂ = 2 Gt C), we calculate a minimum range of inputs of carbon into the atmospheric reservoir by simply assuming no coeval uptake by organic carbon burial, silicate weathering or fertilization of the terrestrial biosphere. According to this scenario, a p_{CO_2} spike ranging from 1,300 to 2,200 p.p.m.v. is equivalent to a 2,600 to 4,400 Gt C input into the atmospheric reservoir.

Received 2 November 2004; accepted 24 March 2005.

- Jenkyns, H. C. The Early Toarcian (Jurassic) anoxic event—Stratigraphic, sedimentary, and geochemical evidence. *Am. J. Sci.* **288**, 101–151 (1988).
- Arthur, M. A., Dean, W. E. & Pratt, L. M. Geochemical and climatic effects of increased marine organic carbon burial at the Cenomanian/Turonian boundary. *Nature* **335**, 714–717 (1988).
- Jenkyns, H. C. Evidence for rapid climate change in the Mesozoic-Palaeogene greenhouse world. *Phil. Trans. R. Soc. Lond. A* **361**, 1885–1916 (2003).
- Hesselbo, S. P. et al. Massive dissociation of gas hydrate during a Jurassic oceanic anoxic event. *Nature* **406**, 392–395 (2000).
- Harries, P. J. & Little, C. T. S. The early Toarcian (Early Jurassic) and the Cenomanian-Turonian (Late Cretaceous) mass extinctions: similarities and contrasts. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **154**, 39–66 (1999).
- Kerr, A. C. Oceanic plateau formation; a cause of mass extinction and black shale deposition around the Cenomanian-Turonian boundary? *J. Geol. Soc. Lond.* **155**, 619–626 (1998).
- Kump, L. R. & Arthur, M. A. Interpreting carbon-isotope excursions; carbonates and organic matter. *Chem. Geol.* **161**, 181–198 (1999).
- Svensen, H. et al. Release of methane from a volcanic basin as a mechanism for initial Eocene global warming. *Nature* **429**, 542–545 (2004).
- McElwain, J. C. Do fossil plants signal palaeoatmospheric CO₂ concentration in the geological past? *Phil. Trans. R. Soc. Lond. B* **353**, 83–95 (1998).
- Royer, D. L., Berner, R. A. & Beerling, D. J. Phanerozoic atmospheric CO₂ change: evaluating geochemical and paleobiological approaches. *Earth Sci. Rev.* **54**, 349–392 (2001).
- Koppelhus, E. B. & Nielsen, L. H. Palynostratigraphy and paleoenvironments of the Lower to Middle Jurassic Bagå Formation of Denmark. *Palynology* **18**, 139–194 (1994).
- Michelson, O., Nielsen, L. H., Johannessen, P. N., Andsbjerg, J. & Surlyk, F. in *The Jurassic of Denmark and Greenland* (eds Ineson, J. R. & Surlyk, F.) 147–216 (Geological Survey of Denmark and Greenland Bulletin, Copenhagen, 2003).
- McElwain, J. C., Mayle, F. E. & Beerling, D. J. Stomatal evidence for a decline in atmospheric CO₂ concentration during the Younger Dryas stadial: a comparison with Antarctic ice core records. *J. Quat. Sci.* **17**, 21–29 (2002).
- Kürschner, W. M. The anatomical diversity of recent and fossil leaves of the durmast oak (*Quercus petraea* Lieblein *Quercus pseudocastanea* Goeppert): implications for their use as biosensors of palaeoatmospheric CO₂ levels. *Rev. Palaeobot. Palynol.* **96**, 1–30 (1997).
- Palfy, J. & Smith, P. L. Synchrony between Early Jurassic extinction, oceanic anoxic event, and the Karoo-Ferrar flood basalt volcanism. *Geology* **28**, 747–750 (2000).
- Cleveland, W. S. & Devlin, S. J. Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**, 596–610 (1988).
- Royer, D. L. et al. Paleobotanical evidence for near present-day levels of atmospheric CO₂ during part of the Tertiary. *Science* **292**, 2310–2313 (2001).
- Jenkyns, H. C., Jones, C. E., Gröcke, D. R., Hesselbo, S. P. & Parkinson, D. N. Chemostratigraphy of the Jurassic System: applications, limitations and implications for palaeoceanography. *J. Geol. Soc. Lond.* **159**, 351–378 (2002).
- Bailey, T. R., Rosenthal, Y., McArthur, J. M., van de Schootbrugge, B. & Thirlwall, M. F. Paleocyanographic changes of the Late Pleinsbachian-Early Toarcian interval: a possible link to genesis of an oceanic anoxic event. *Earth Planet. Sci. Lett.* **212**, 307–320 (2003).
- Vakhrameev, V. A. *Jurassic and Cretaceous Floras and Climates of the Earth* (Cambridge Univ. Press, Cambridge, 1991).
- Cohen, A. S., Coe, A. L., Harding, S. M. & Schwark, L. Osmium isotope evidence for the regulation of atmospheric CO₂ by continental weathering. *Geology* **32**, 157–160 (2004).
- Siegenthaler, U. & Sarmiento, J. L. Atmospheric carbon dioxide and the oceans. *Nature* **365**, 119–125 (1993).
- Milkov, A. V. Global estimates of hydrate-bound gas in marine sediments: how much is really out there? *Earth Sci. Rev.* **66**, 183–197 (2004).
- Veevers, J. J., Powell, C. M., Collinson, J. W. & Lopez-Gamundi, O. R. in *Permian-Triassic Pangean Basins and Foldbelts along the Panthalassan Margin of Gondwanaland* (eds Veevers, J. J. & Powell, C. M.) 331–353 (Geological Society of America Memoir, Boulder, 1994).
- Elliot, D. H. & Fleming, T. H. Occurrence and dispersal of magmas in the Jurassic Ferrar Large Igneous Province, Antarctica. *Gondwana Res.* **7**, 223–237 (2004).
- Ballance, P. F. & Watters, W. A. Hydrothermal alteration, contact metamorphism, and authigenesis in Ferrar Supergroup and Beacon Supergroup rocks, Carapace Nunatak, Allan Hills, and Coombs Hills, Victoria Land, Antarctica. *NZ J. Geol. Geophys.* **45**, 71–84 (2002).
- Kurtz, A. C., Kump, L. R., Arthur, M. A., Zachos, J. C. & Paytan, A. Early Cenozoic decoupling of the global carbon and sulfur cycles. *Paleoceanography* **18**, 1090–1104 (2003).
- Knoll, A. H., Bambach, R. K., Canfield, D. E. & Grotzinger, J. P. Comparative Earth history and Late Permian mass extinction. *Science* **273**, 452–457 (1996).
- Beerling, D. J., Lomas, M. R. & Gröcke, D. R. On the nature of methane gas-hydrate dissociation during the Toarcian and Aptian oceanic anoxic events. *Am. J. Sci.* **302**, 28–49 (2002).
- Kothavala, Z., Oglesby, R. J. & Saltzman, B. Sensitivity of equilibrium surface temperature of CCM3 to systematic changes in atmospheric CO₂. *Geophys. Res. Lett.* **26**, 209–212 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Comer Foundation of Science and Education for funding, F. Surlyk for logistical support for fieldwork, I. Glasspool for insights on Karoo basin coal, P. Wagner for statistical assistance, and M. Huber and B. Sageman for critical comments on an earlier version of the manuscript. Scientific discussion with M. Arthur, H.C. Jenkyns and L.R. Kump is also acknowledged. We thank B.S. Cramer for suggestions that greatly improved the manuscript.

Author Contributions J.C.M. and J.W.M. were responsible for laboratory data collection. All authors contributed equally to interpretation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.C.M. (jmcclwain@fieldmuseum.org).

Seismological evidence for mosaic structure of the surface of the Earth's inner core

Dmitry N. Krasnoshchekov¹, Peter B. Kaazik¹ & Vladimir M. Ovtchinnikov¹

The transition from the Earth's solid inner core to liquid outer core is the location where the inner core grows¹ and from which compositional convection in the outer core originates^{2,3}. Most seismological models of the Earth describe the inner-core boundary as sharp^{4,5} and simple^{6,7,8}, although experimental data requiring the presence of a thin transition layer at the bottom of the outer core have been reported⁹. The density jump at the inner-core boundary—an important parameter determining gravitational energy release¹⁰ and constraining the compositional difference between the inner and outer core—is also not well known. Estimates of this density jump obtained using free-oscillation eigenfrequencies give low values^{11–13} of 0.25–1.0 g cm^{−3}, whereas a method using the amplitude ratio of core-reflected phases yielded values of 0.6–1.8 g cm^{−3} (refs 14–17). Here we analyse properties of waves precritically reflected from the Earth's inner core (PKiKP phases) that show significant variability in amplitude, consistent high-frequency content and stable travel times with respect to a standard Earth model⁴. We infer that the data are best explained by a mosaic structure of the inner core's surface. Such a mosaic may be composed of patches in which the transition from solid inner to liquid outer core includes a thin partially liquid layer interspersed with patches containing a sharp transition.

Precritical PKiKP waveforms are sensitive to the composition and fine structure of the inner-core boundary (ICB). Being a weak phase frequently obscured by seismic coda, PKiKP arrivals are hard to detect and thought to be anomalously large¹⁶ if observed (for example due to focusing), or require array data processing for unambiguous identification (such as linear or phase-weighted stacking of vertical channels^{15,18} or velocity filtering⁷). PKiKP observations have not become routine and are sometimes either extremely unusual, owing to the surprisingly low magnitude of seismic sources¹⁹ (body-wave magnitude, $M_b < 5.0$), or questionable, because of significant PKiKP travel-time residuals with respect to standard models (up to 12 s)¹⁵. The repeatedly investigated^{16,19} question of PKiKP observability has sometimes been explained by claiming that PKiKP observations are due to differences in the source radiation pattern^{15,17}, which is not entirely convincing. The ICB density-jump estimates derived from the PKiKP/PcP ratio are often questioned because of considerable PKiKP and PcP amplitude variation⁷ and the low reliability of non-array data or single PKiKP observations (where a pulse from a random process may be interpreted as a PKiKP waveform). Here we revisit the issue of the fine structure of the ICB and its reflection properties, using a new reliable data set of precritical PKiKP waveforms observed between 6° and 90° following underground nuclear explosions (UNEs) in the USSR, USA and China (see 'Data selection' in Methods).

Precritical PKiKP waveforms on a single vertical channel are basically hidden by the intensive oscillations of seismic coda, making

it hard to obtain a good signal-to-noise ratio. However, the ratio can be effectively improved by band-pass filtering; this revealed, on vertical components, waveforms dominating in the 40-s time interval around the PKiKP arrival time predicted by PREM (Fig. 1). Although the resulting signal-to-noise ratio on single filtered traces varies from 1.2 to 3.5, the PKiKP waveform is recognizable on each individual channel by eye. No such waveforms were observed on horizontal components. The 31 PKiKP pulses observed at BRVK (see Fig. 1) correspond to explosions with $M_b > 5.7$, whereas no PKiKP waveforms were registered for UNEs with $M_b \leq 5.6$, no matter how sensitive the registration seismic channel was (up to 0.06 nm per count). Phase stacking of vertical components for 12 such weaker UNEs also revealed no PKiKP waveform on the sum trace. The absence of PKiKP waveforms from records of UNEs with $M_b < 5.7$ is also noted when processing data from other stations (such as NRN and FRU), which may point to an experimental PKiKP observability magnitude threshold where data from isotropic explosive sources are used.

The detected PKiKP arrivals have stable travel times consistent with PREM, with residuals not exceeding 1.5 s (70% are within 0.5 s). The observed PKiKP waveforms feature good correlation and robust dynamic properties for each site–station pair. The PKiKP statistics for 31 Semipalatinsk explosions with magnitudes 5.8–6.1 registered at BRVK are as follows: the amplitude (reduced to $M_b = 5.9$) is 3.30 ± 0.93 nm (see 'Reduction of amplitudes' in Methods), the period is 0.66 ± 0.06 s, and travel time is 992.45 ± 0.07 s (PREM theoretical travel time is 992.42 s). We particularly note the almost equal PKiKP amplitudes generated by the same Semipalatinsk UNEs recorded at 6° (BRVK) and 89° (NWA0). Although the PKiKP amplitude predicted by PREM for 89° is at least 30 times smaller than for 6°, the measured PKiKP amplitudes at NWA0 were 2.65 ± 0.40 nm. The PKiKP periods measured at NWA0 were all 0.67 s.

Another essential and common feature of all the detected PKiKP waveforms is their high-frequency content, with typical periods 0.6–0.83 s (75% are 0.7 s). Unlike P waves, the frequency content of the PKiKP waveforms remains unchanged for all events with magnitudes from 5.7 to 7.0: the measured PKiKP periods generated by nine explosions with $6.5 < M_b < 7.0$ were 0.73 ± 0.06 s. In contrast to P waves, which have a set of equally powerful frequency maxima, all PKiKP spectrograms show a single accentuated frequency (see Supplementary Fig. 2), with other source frequency maxima damped.

There were also cases in which we failed to detect distinct PKiKP arrivals using band-pass filtering or linear and phase-weighted stacking, for reasons that are uncertain. Application of different narrow- and wide-frequency band filters in the range 0.5–5 Hz sometimes did not improve the PKiKP signal-to-noise ratio (Fig. 2). As no dramatic deviations are registered in the P and PcP

¹Institute for Dynamics of the Geospheres, Russian Academy of Sciences, Leninsky pr. 38, korp. 1, Moscow, 119334, Russia.

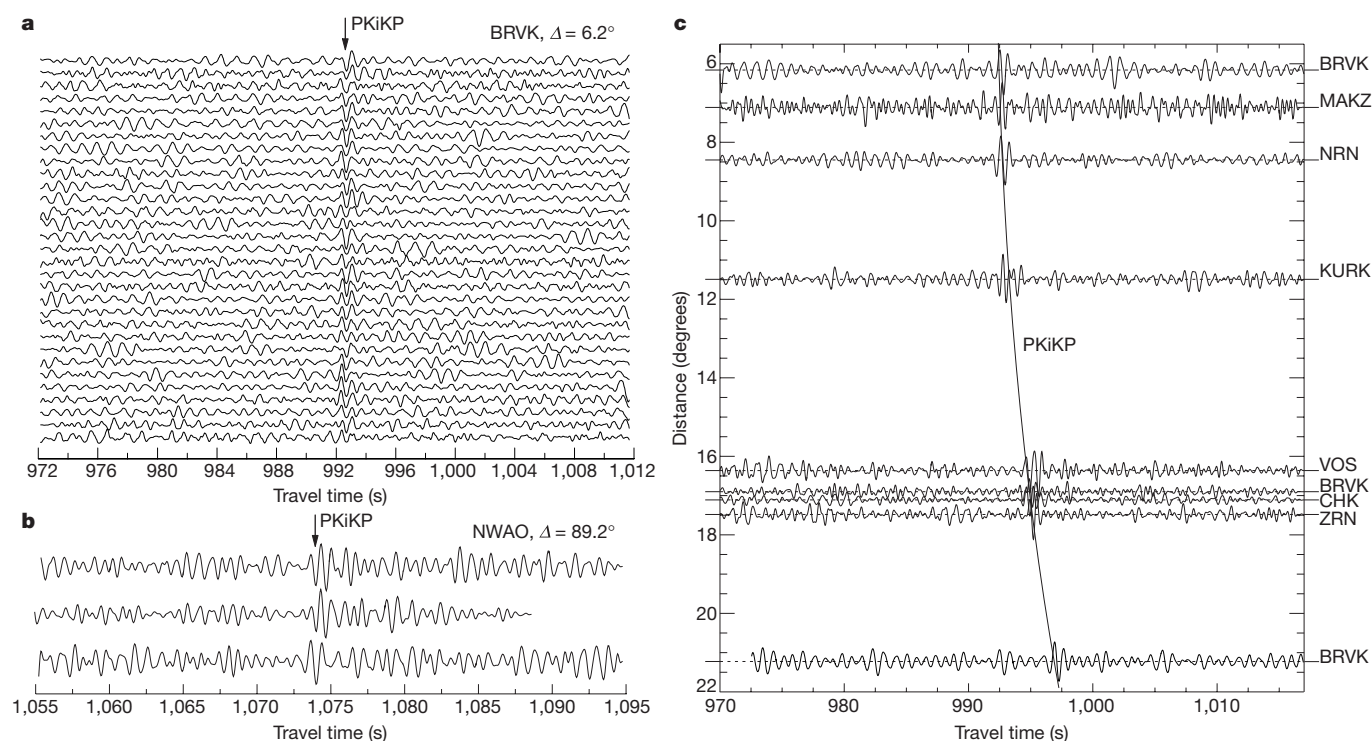


Figure 1 | Band-pass-filtered vertical components of UNE records and predicted travel times. Traces correspond to 31 Semipalatinsk explosions recorded by station BRVK in Kazakhstan (a), three Semipalatinsk explosions recorded by station NWAQ in Australia (b), and Semipalatinsk, Lop Nor and Novaya Zemlya explosions recorded by stations in Central Asia (c). PKiKP

theoretical travel-time curve and arrivals are computed with respect to PREM. All traces are filtered between 1.0 and 5.0 Hz with a three-pole zero-phase Butterworth filter. Δ , distance. Locations of the stations are shown in Supplementary Fig. 1.

phases on the appropriate fragments of the NRN record, and PKiKP waveforms generated by this same explosion are clearly detected on three-component records from other stations, the influence of the seismic source can be excluded as a factor causing such non-observation. Meanwhile, PKiKP pulses are detected by band-pass filtering of NRN records after weaker UNEs performed elsewhere (Semipalatinsk and PNE (peaceful nuclear explosion) examples in

Fig. 2), which is evidence for PKiKP observability despite any possible unfavourable local station factors or local geophysical conditions that might prevent such observations.

The PKiKP amplitudes measured at NRN after the Semipalatinsk explosions and PNEs varied between 2 and 6 nm, which yields a PKiKP amplitude estimate of 22 nm for 31.38° and $M_b = 7.0$. However, despite the quite low root-mean-square (r.m.s.) coda

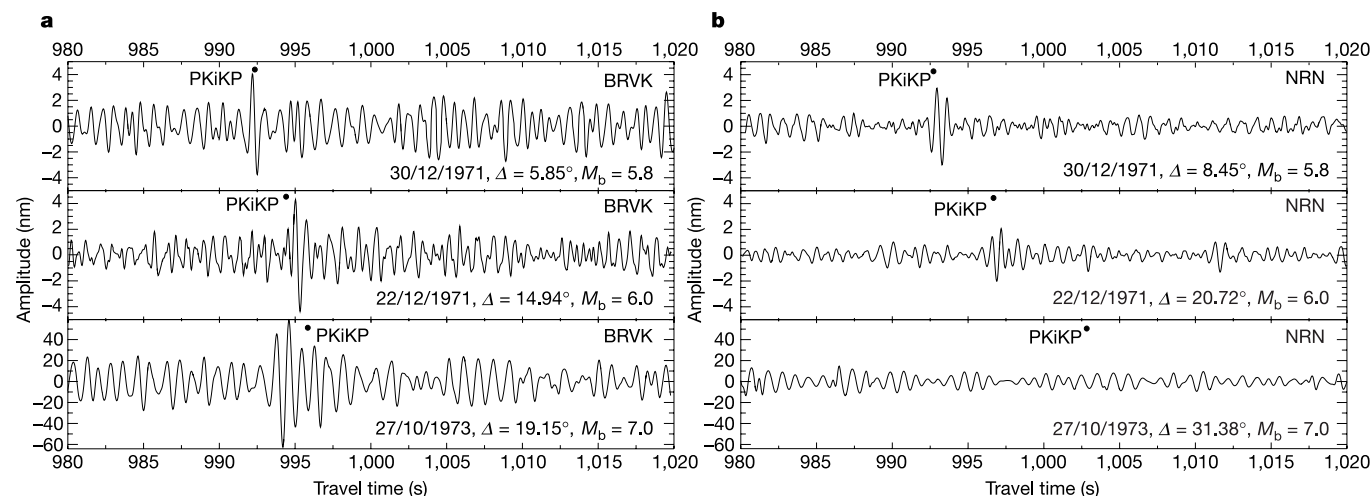


Figure 2 | Band-pass-filtered vertical components of three explosions measured at two stations. a, BRVK station; b, NRN station. The traces correspond to the Semipalatinsk explosion dated 30/12/1971, the PNE (peaceful nuclear explosion, carried out for industrial, non-military purposes) dated 22/12/1971 and one Novaya Zemlya explosion dated 27/10/1973. Dots

indicate theoretical PKiKP arrival times computed with respect to PREM without corrections for the Earth's ellipticity. The correction would decrease the PKiKP theoretical travel time for the Novaya Zemlya explosion, making the prediction closer to the observed arrival.

level of 4.7 nm on a filtered vertical component in the vicinity of the expected arrivals, we could not identify a PKiKP waveform on three-component NRN records of available Novaya Zemlya explosions. The close theoretical arrivals of seismic coda for 2 min before and 30 s after the PKiKP include the Lg and ScS phases. The Lg arrival, depending on the Earth model used, is 6–12 s before PKiKP, while ScS is predicted to travel 6–15 s longer than PKiKP. Propagation of Lg waves is blocked by a marine passage of the seismic path,

and we would not expect anomalously early arrivals of ScS to contaminate our PKiKP window because of the use of explosions. Besides, ScS has longer periods and mainly horizontal particle motion. Thus, it is doubtful that PKiKP arrivals are obscured by a more powerful phase or seismic coda.

To estimate the PKiKP amplitudes not detected by band-pass filtering, we used stacking techniques and performed frequency–wavenumber (f – k) analysis for a group of explosions carried out at a given test site and registered by a station. Unlike PcP, the PKiKP pulse is clearly seen only on the sum trace and is usually not tracked on each individual record (Fig. 3). The resulting signal-to-noise ratio in adjacent 5-s windows of the beam is 2.1, and the estimated P–PKiKP slowness is $0.062 \pm 0.012 \text{ s km}^{-1}$ (0.051 s km^{-1} according to PREM). The estimated PKiKP r.m.s. amplitude corresponding to the slowness determined in Fig. 3c is $1.40 \pm 0.21 \text{ nm}$. Misidentification of seismic phases from the diagram is ruled out, as all nearby phases have significantly higher slownesses attenuated by the wave-number filter. In contrast to the non-observation of PKiKP at 31.38° on NRN records, this velocity filtering was effective when detecting PKiKP on records from station KEV located 31.4° away from the Semipalatinsk array of explosions. The resulting estimate of the PKiKP amplitude arriving closely after Lg with its low phase velocity of 3.5 km s^{-1} yielded $4.31 \pm 1.30 \text{ nm}$.

Figure 4 features 59 direct body-wave measurements of PKiKP amplitudes, four source-array estimates (see Supplementary Table), and recently published²⁰ PKiKP amplitude evaluation for 89.5° . We note a substantial variability of the observed PKiKP amplitudes, giving a spatially complicated pattern. At distances up to 35° , the scatter of PKiKP amplitudes (reduced to $M_b = 5.9$) is a factor of six, while amplitude variability before reduction is expected to be only a factor of two to three⁷ for explosions with magnitudes 5.8–6.1. Whereas measured PKiKP travel times and amplitudes can be simulated within models with sharp ICB by varying the ICB density jump and outermost inner core's S-wave velocity, the form of the detected arrivals requires the introduction of either specific attenuation patterns or a transition layer (see 'Modelling' in Methods). To account for individual PKiKP waveforms, trial amplitude curves calculated with respect to PREM and its modifications with a liquid/solid transition layer were evaluated. Thus, in Fig. 4, the liquid layer's parameters were selected so as to match PKiKP pulses observed at 6° using the source function derived from modelling of P. The liquid layer was chosen because the observed amplitudes do not decline with distance as predicted by standard and solid-layer models. However, none of the trial curves shows good agreement with experimental data (see PKiKP amplitudes in the so-called transparent zone 50 – 70° , or the tenfold discrepancy between our PKiKP body-wave measurement at 89.2° and the reduced-amplitude estimate²⁰ at 89.5° where PREM theoretical amplitude is negligible).

The cause of PKiKP amplitude variability and high-frequency content is unlikely to be associated with shallow structures or the core–mantle boundary's strongly heterogeneous region. The former scenario would result in PcP and PKiKP phases being affected in a similar way, while the latter would inevitably lead to considerable

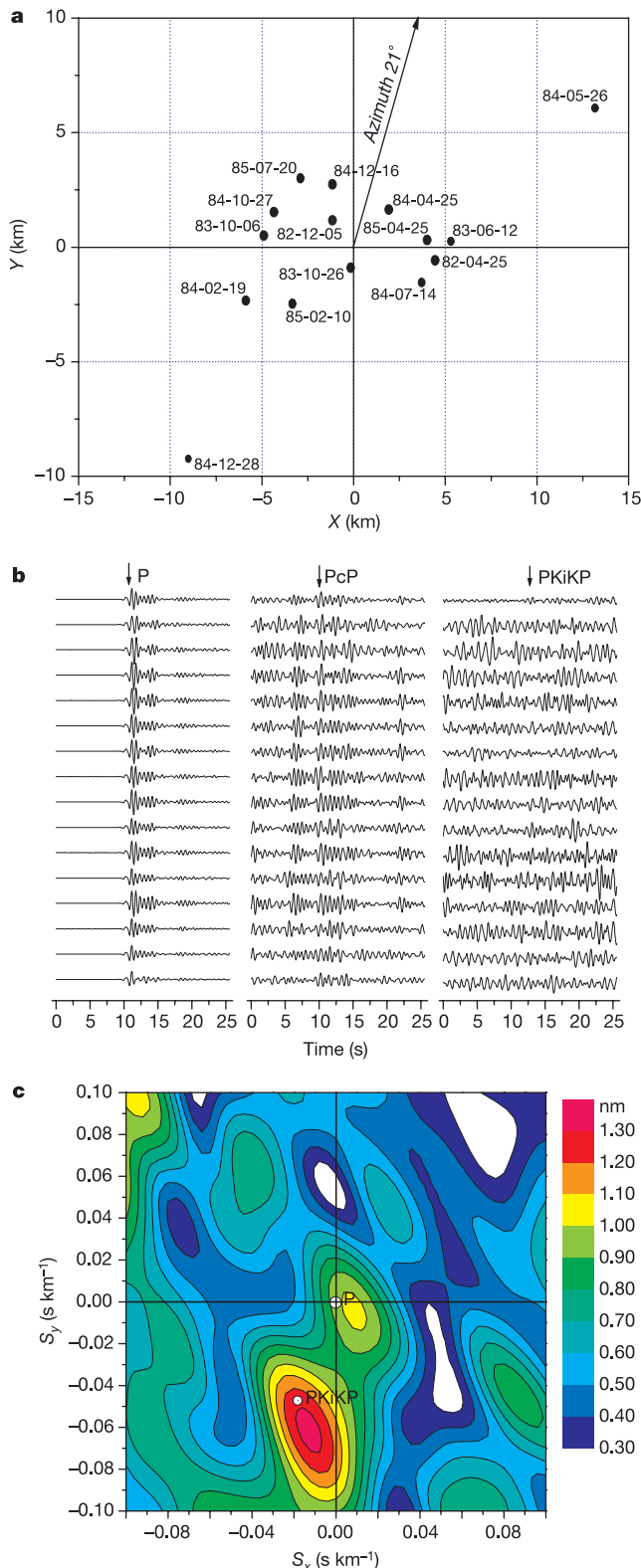


Figure 3 | Seismic body-wave data used for f – k analysis and the results. **a**, Array of 15 Semipalatinsk explosions recorded at COL, and azimuth to the station; explosion dates are in the yy-mm-dd format. **b**, Fragments of band-pass-filtered records corresponding to arrivals of P, PcP and PKiKP waves, and the sum trace (top); the PcP arrivals were used to check introduced corrections for different magnitudes of the explosions. **c**, Slowness diagram calculated in the 1.5-s time window corresponding to the arrival of the PKiKP phase; as records are aligned on first arrival of the P wave, the resulting slowness diagram is relative to P; the white dot corresponds to the theoretical estimate of the PKiKP slowness and azimuth calculated with respect to PREM and geographical locations. S_x and S_y represent the magnitudes of slowness vectors s_x and s_y .

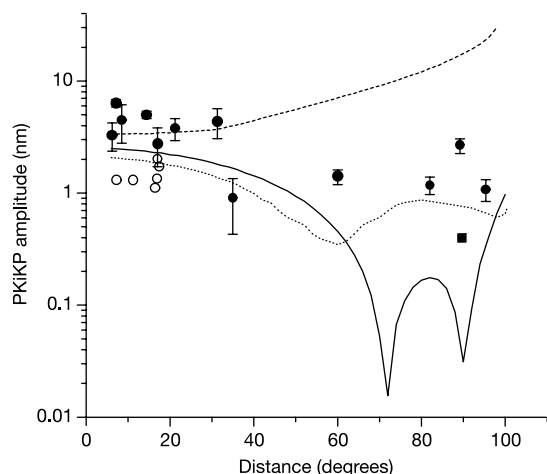


Figure 4 | Measured PKiKP amplitudes and theoretical amplitude curves. Open circles, single measurements; filled circles, averaged reduced amplitudes generated by test site's explosions recorded by a station. The square is the PKiKP amplitude estimate from ref. 20. The solid line denotes the PKiKP amplitude curve calculated with respect to standard PREM; dashed line, PREM modification with ICB liquid transition layer; dotted line, PREM modification with ICB solid transition layer. The density of 2.2-km-thick layers corresponds to the bottom of the outer core ($12.1663 \text{ g cm}^{-3}$) and the top of the inner core ($12.7636 \text{ g cm}^{-3}$) for liquid and solid layers respectively, while P-wave velocity is 12 km s^{-1} . PKiKP amplitude errors are ± 1 standard deviation defined after simple averaging or bootstrap resampling as referenced in Methods.

PKiKP travel-time anomalies exceeding the observed 1.5 s. In addition, such focusing D'' structure can hardly provide amplifications exceeding²¹ 50%. Local ICB complexities such as topography or lateral dependency of the ICB reflection coefficient due to changes in ICB fine structure and/or composition seem more credible and are in line with recent ICB studies appealing to reverberations²² instead of focusing/scattering effects. The ICB topography recently limited¹⁹ by a peak-to-peak amplitude of 4 km can hardly provide the required effects. In addition, any pronounced local topography must comply with appropriate viscosity constraints²³ and inner-core differential rotation²⁴. Rather, specific local surface conditions at the PKiKP reflection point are likely to control the PKiKP properties. Just like mosaic patches, the regions of homogeneous reflection coefficient can be defined by uniform ICB fine structure (sharp transition or with a layer) and physical composition. Patches with a liquid transition layer amplify high-frequency PKiKP arrivals, while reflections from sharp discontinuity patches result in weak or no PKiKP observations unless the focal mechanism is favourable and the body-wave magnitude of the event is big enough to generate a strong observable PKiKP phase. It takes slight variations in the patches' properties to account for our PKiKP data—up to 5 km in thickness, and up to 8% P-wave velocity jump above the inner-core velocity, which is within the 10% uncertainty²⁵ for the latter parameter.

Estimates of characteristic lateral dimensions of mosaic patches would require a great number of reflection points probing the inner core's surface. However, the lower bound can be assessed on the basis of the body wave's seismic velocity and observed PKiKP wavelengths, which yields a figure of 10 km. The distribution of available reflection points through the inner-core surface beneath Eurasia shows significant changes in the character of the reflections when the reflection points are separated by distances of 23–240 km. Meanwhile, our data set is quite sparse and has limited geographical coverage, precluding any bigger regions characterized by the presence of a liquid layer in the transition.

Non-uniform boundary freezing and compositional convection

are the best possible mechanisms to explain the inner core's surface patchiness. The mosaic patches could be formed by vast regions with similar temperature regimes or by buoyant liquid blobs released from the ICB mushy zone³. Considering the changeable nature of the reflections, we favour the second, more local, approach, with patches from tens to hundreds of kilometres.

METHODS

Data selection. The use of UNE records eliminates the necessity of allowing for peculiarities due to different radiation patterns of seismic events. Explosive sources are also much simpler than earthquakes and characterized by a lower intensity of S waves. By using non-seismological (ground truth) precise information on explosive source parameters we exclude errors caused by uncertainties in locations, times and energy of seismic sources. Our data set includes more than 200 short-period and broadband digital seismic records from 23 stations located in Eurasia, Australia and North America (see Supplementary Fig. 1). The sensitivity of selected seismic channels meets the requirement of being able to provide undistorted registering of at least a few tenths of nanometres in the seismic waveform frequency range from 0.5 to 5 Hz. We selected stations that recorded three or more explosions carried out at the same test site, or records of an explosion simultaneously registered by a number of stations or an array. We consider multiple observations as one of the most important factors contributing to the reliability of the PKiKP detection, and have disregarded any possible PKiKP waveforms appearing on a single record from a station that registered only one explosion.

Reduction of amplitudes. Using the classical definition of a body-wave magnitude as the decimal logarithm of a phase's amplitude-to-period ratio, the measured PKiKP amplitudes are reduced to a magnitude of 5.9, corresponding to the majority of the processed explosions. Low variation in observed PKiKP periods gives another reason for making use of the definition to reduce PKiKP amplitudes from explosions of different magnitudes carried out at the same test site. Thus, the expression $A_{\text{reduced}} = A_{\text{measured}} \times 10^{(5.9-m)}$ is used, where m is the body-wave magnitude of the explosion whose PKiKP amplitude is being reduced. The resulting average of the reduced PKiKP amplitudes generated by the 31 Semipalatinsk explosions with magnitudes between 5.8 and 6.1 is $3.30 \pm 0.93 \text{ nm}$, compared with $3.96 \pm 1.53 \text{ nm}$ before reduction, which decreases the scatter by 39%.

Frequency–wavenumber (f – k) analysis. Stacking array data are proved as an effective means for extracting weak signals, successfully applied to the problem of PKiKP detection^{18,19}. Although summation of seismic records performed using certain methods may lead to some distortion of sought amplitudes, the appropriate estimates have also been confirmed as quite reliable^{26,27}. We had enough data for four stations that recorded an array of powerful Semipalatinsk explosions carried out 30° to 95° away. In this case, test-site dimensions (about 30 km) are small compared to the epicentral distances, which for the purposes of inversion enables us to consider the array of sources as a small aperture array, and hence seismic waves as plane waves. Application of stacking and methods of f – k analysis to seismic data are frequently affected by errors in timing, locations and different magnitudes of events. To perform the analysis independently of a UNE's origin times and possible errors of station clocks, band-pass-filtered traces were aligned on the first arrival of a P wave by means of a cross-correlation technique. A Butterworth three-pole zero-phase filter in the frequency band 1–3 Hz was used for the filtering. The P phase is also used to adjust for different source magnitudes by means of r.m.s. amplitudes of individual traces. Having performed this preliminary processing, we initiate a recursive grid search over two-dimensional, cartesian slowness vectors s_x and s_y to find the combination giving the maximum beam power in a predefined time window. The first search is run with bounds -0.1 to 0.1 s km^{-1} using an increment of 0.004 s km^{-1} for both slowness components. The beam power is defined as r.m.s. amplitude in a 1.5-s time window with linear beams. To calculate error bounds for the array estimates, a bootstrap-type resampling procedure²⁸ was used. Following ref. 18, we used generated pseudoarrays with replacement of records to calculate the samples.

Modelling. To calculate PKiKP synthetic seismograms, we first derive the source function from full-wave modelling²⁹ of the first arrival of P (if detected, the PcP waveform is modelled too), using the explosive source model³⁰. We then proceed by trial and error to select ICB parameters/structure providing the best fit of the observed PKiKP waveform with correlation coefficient exceeding 0.9 (see Supplementary Fig. 3). The modelling error is well below PKiKP amplitude error bounds, which are between 15% and 37%.

Received 24 December 2004; accepted 31 March 2005.

1. Jacobs, J. A. The Earth's inner core. *Nature* **172**, 297–298 (1953).

2. McFadden, P. L. & Merrill, R. T. Geodynamo source constraints from paleomagnetic data. *Phys. Earth Planet. Inter.* **43**, 22–33 (1986).
3. Moffatt, H. K. & Loper, D. E. The magnetostrophic rise of a buoyant parcel in the Earth's core. *Geophys. J. Int.* **117**, 394–402 (1994).
4. Dziewonski, A. M. & Anderson, D. L. Preliminary reference earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
5. Kennett, B. L. N., Engdahl, E. R. & Buland, R. Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.* **122**, 108–124 (1995).
6. Engdahl, E. R., Flinn, E. A. & Masse, R. P. Differential PKiKP travel times and the radius of the inner core. *Geophys. J.R. Astron. Soc.* **39**, 457–463 (1974).
7. Buchbinder, G. G. R., Wright, C. & Poupinet, G. Observations of PKiKP at distances less than 110°. *Bull. Seismol. Soc. Am.* **63**, 1699–1707 (1973).
8. Cummins, P. & Johnson, L. Synthetic seismograms for an inner core transition of finite thickness. *Geophys. J.* **94**, 21–34 (1988).
9. Adushkin, V., An, V., Ovtchinnikov, V. & Krasnoshchekov, D. On the density jump at the inner boundary of the Earth's core from observations of PKiKP waves at distances of about 6°. *Dokl. Russ. Acad. Sci./Earth Sci.* **334**, 595–598 (1997).
10. Gubbins, D., Masters, T. & Jacobs, J. Thermal evolution of the Earth's core. *Geophys. J.R. Astron. Soc.* **59**, 57–99 (1979).
11. Gilbert, F. & Dziewonski, A. An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra. *Phil. Trans. R. Soc. Lond. A* **278**, 187–269 (1975).
12. Masters, T. G. Observational constraints on the chemical and thermal structure of the Earth's deep interior. *Geophys. J.R. Astron. Soc.* **57**, 507–534 (1979).
13. Masters, G. & Gubbins, D. On the resolution of density within the Earth. *Phys. Earth Planet. Inter.* **140**, 159–167 (2003).
14. Bolt, B. A. & Qamar, A. An upper bound to the density jump at the boundary of the Earth's inner core. *Nature* **228**, 148–150 (1970).
15. Souriau, A. & Souriau, M. Ellipticity and density at the inner core boundary from subcritical PKiKP and PcP data. *Geophys. J. Int.* **98**, 39–54 (1989).
16. Shearer, P. & Masters, G. The density and shear velocity contrast at the inner core boundary. *Geophys. J. Int.* **102**, 491–498 (1990).
17. Cao, A. & Romanowicz, B. Constraints on density and shear velocity contrasts at the inner core boundary. *Geophys. J. Int.* **157**, 1146–1151 (2004).
18. Koper, D. K. & Pyle, M. L. Observations of PKiKP/PcP amplitude ratios and implications for Earth structure at the boundaries of the liquid core. *J. Geophys. Res.* **109**, B03301, doi:10.1029/2003JB002750 (2004).
19. Koper, D. K., Pyle, M. L. & Franks, J. M. Constraints on aspherical core structure from PKiKP-PcP differential travel times. *J. Geophys. Res.* **108**(B3), 2168, doi:10.1029/2002JB001995 (2003).
20. Adushkin, V. V. & Ovtchinnikov, V. M. The mosaic in reflections from the Earth's solid core boundary. *Dokl. Earth Sci. A* **397**(6), 883–885 (2004).
21. Tibuleac, I. M., Nolet, G., Michaelson, C. & Koulakov, I. P wave amplitudes in a 3-D earth. *Geophys. J. Int.* **155**, 1–10 (2003).
22. Poupinet, G. & Kennett, B. L. N. On the observation of high frequency PKiKP and its coda in Australia. *Phys. Earth Planet. Inter.* **146**, 497–511 (2004).
23. Buffet, B. A. Geodynamic estimates of the viscosity of the Earth's inner core. *Nature* **388**, 571–573 (1997).
24. Song, X. D. & Richards, P. G. Seismological evidence for differential rotation of the Earth's inner core. *Nature* **392**, 221–224 (1996).
25. Song, X. & Helmberger, D. V. A P-wave velocity model of the Earth's core. *J. Geophys. Res.* **100**(B7), 9817–9830, doi:10.1029/94JB03135 (1995).
26. Schimmel, M. & Paulssen, H. Noise reduction and detection of weak, coherent signals through phase-weighted stacks. *Geophys. J. Int.* **130**, 497–505 (1997).
27. McFadden, P. L., Drummond, B. J. & Kravis, S. The Nth-root stack: Theory, applications, and examples. *Geophysics* **51**, 1879–1892 (1986).
28. Efron, B. & Tibshirani, R. Statistical data analysis in the computer age. *Science* **253**, 390–395 (1991).
29. Richards, P. G. Calculation of body waves, for caustics and tunnelling in core phases. *Geophys. J.R. Astron. Soc.* **35**, 243–264 (1973).
30. Muller, G. Seismic moment and long-period radiation of underground nuclear explosions. *Bull. Seismol. Soc. Am.* **63**, 847–857 (1973).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The research described was made possible in part by contribution from the US Civilian Research & Development Foundation for the Independent States of the Former Soviet Union (CRDF) and individual financial support provided by the Russian Science Support Foundation (RSSF) to D.K.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.K. (krasnd@idg.chph.ras.ru).

LETTERS

Tree use by koalas in a chemically complex landscape

Ben D. Moore^{1†} & William J. Foley¹

Although defence against herbivores is often argued to be the main action of plant secondary metabolites (PSMs)¹, very few examples have demonstrated that intraspecific variation in PSM concentrations influences foraging by wild vertebrate herbivores^{2,3}. Experiments with captive animals often indicate that PSM concentrations influence how much herbivores eat from individual plants^{3–7}, but these experiments do not replicate the subtle trade-offs in diet selection faced by wild animals, which must avoid predators and extremes of weather, interact with conspecifics, and achieve a balanced, nutritious diet, while avoiding intoxication by PSMs. We characterized the foliar chemistry of every tree from two *Eucalyptus* species available to a population of koalas (*Phascolarctos cinereus*) and considered rates of tree visitation over a ten-year period. We show that visitation rate was most strongly influenced by tree size, but that koalas also visited trees less frequently if the foliage contained either high concentrations of deterrent PSMs known as formylated phloroglucinol compounds, or low concentrations of nitrogen. Consequently, plant chemistry restricts the use of trees by this herbivore, and thus limits the food available to koalas and potentially influences koala populations.

Although feeding experiments with captive animals often show that the intake of foliage by herbivores is a linear function of PSM concentrations^{4,6,8}, more-complex relationships may exist between levels of plant defence and rates of herbivory by wild animals. For example, high PSM concentrations may limit the potential nutritional value of a plant by restricting herbivores' intake, and detoxification of PSMs may entail time, energetic or nutritional costs^{9,10}. Beyond the point at which the net gain from eating a plant is exceeded by the PSM-related costs, herbivory may be prevented entirely, but herbivores may be less discriminating among plants with PSM concentrations below such a threshold. Furthermore, wild herbivores may counter the effects of plant toxins by eating mixed diets, thereby distributing detoxification costs across multiple detoxification pathways¹⁰ or by selecting nutrient-rich foliage^{9,11–13}.

Koalas are highly specialized folivores of *Eucalyptus*, primarily of the subgenus *Symphyomyrtus*¹⁴, in which a class of lipophilic phenolic compounds known as formylated phloroglucinol compounds (FPCs) are widespread¹⁵. Individual *Eucalyptus* trees are large and long-lived, and so represent clearly defined feeding patches¹⁶ in which the presence of a koala can be clearly and unambiguously observed. We considered tree use by koalas in a *Eucalyptus* woodland dominated by three species: *Eucalyptus globulus* (38% of *Eucalyptus* trees), *E. viminalis* (29%) and *E. ovata* (32%), each of which display intraspecific variation in FPC concentrations¹⁷. The amount of *E. globulus* and *E. viminalis* foliage that koalas eat in captivity is inversely related to FPC concentration, but concentrations in *E. ovata* never reach levels sufficient to deter koalas from feeding⁸, so we did not consider that species in this study. Only 16% of 1,264

observations of koala use of *Eucalyptus* trees were in *E. ovata*, so this species accounts for only a small proportion of tree use in this woodland.

We described two key aspects of the foliar chemistry of all 857 *E. globulus* and *E. viminalis* trees present within the 7.6-hectare *Eucalyptus* woodland: total FPC concentration (indicating defence) and total nitrogen concentration (indicating nutritional quality) of the foliage. *Eucalyptus* contains extremely low nitrogen concentrations¹⁸ and many studies have concluded that nitrogen is of particular importance to the koala^{19–23}. Regression analysis showed no significant relationship between FPC and nitrogen concentrations in *E. globulus* and only a very weak, positive relationship in *E. viminalis* ($r^2 = 0.019$, $P < 0.01$).

Having characterized the foliar chemistry of the entire populations of both study species, we adopted a new approach to investigating the relationship between tree attributes (both chemical and morphological) and koala visitation rates. Conventional regression techniques were inappropriate for modelling the relationship between tree visitation and chemical or morphological attributes, because the expected relationship was triangular, rather than linear²⁴. Although we predicted that trees with high FPC concentrations would receive low rates of koala visitation, we expected variable rates of visitation to trees with low FPC concentrations. Even if trees possess low levels of defence, they may be visited only rarely if an unmeasured attribute makes them unattractive to folivores, or simply if the requirements of the folivore population are met by other equally or more suitable trees.

The first attribute we tested was tree size (circumference), because we expected that bigger trees would receive more koala visits, as they represent larger food patches and account for a greater proportion of the foliar biomass available to koalas. We compared the observed distribution of koala visits across trees to the expected distribution if koalas had visited trees randomly and found that koalas used trees that were on average significantly larger than expected (Table 1).

Table 1 | Observed and expected mean tree circumferences

	<i>E. globulus</i>	<i>E. viminalis</i>
n_{obs}	907	421
n_{avail}	484	373
$\bar{\mu}_{\text{obs}}$	205.3	111.8
$\bar{\mu}_{\text{exp}}$	127.8	62.6
95% confidence interval	121.9–133.7	57.4–68.0
$P(\bar{\mu}_{\text{obs}} \leq \bar{\mu}_{\text{exp}})$	< 0.0001	< 0.0001

Mean tree circumference ($\bar{\mu}_{\text{obs}}$, in centimetres) was calculated across all n_{obs} observed koala visits and compared to the expected mean circumference ($\bar{\mu}_{\text{exp}}$). $\bar{\mu}_{\text{exp}}$ and its 95% confidence interval were estimated by randomly selecting n_{obs} values (with replacement) from the list of n_{avail} available trees.

¹School of Botany and Zoology, Australian National University, Canberra ACT 0200, Australia. [†]Present address: School of Tropical Biology, James Cook University, Townsville QLD 4811, Australia.

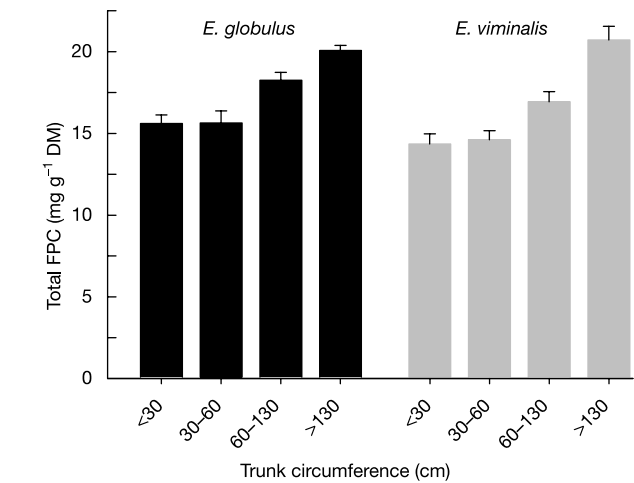


Figure 1 | FPC concentrations in four tree size classes. Mean FPC concentrations (with one standard error) in each of four size classes of tree, for *E. globulus* (black bars) and *E. viminalis* (grey). DM, dry matter basis.

In both species, FPC concentration was positively correlated with tree size (Supplementary Information, also Fig. 1), so by biasing their visits towards larger-than-average trees, koalas were limiting their dietary choices to a subset of trees with higher-than-average FPC concentrations. Therefore, we designed a randomization test to determine whether foliar chemistry influenced visitation rates once the distribution of visits with respect to tree size had been accounted for. We found that mean FPC concentrations were significantly lower and mean nitrogen concentrations were significantly greater than expected among trees visited by koalas (Table 2). However, the observed cumulative distributions of visits across FPC concentrations (Fig. 2) deviated most strongly from the expected distributions at high FPC concentrations, demonstrating that FPCs were only effective in reducing koalas' use of the most highly defended trees. In contrast, trees with moderately high FPC concentrations were used more than expected, and trees with median-to-low FPC concentrations were visited at the rate expected. Similarly, koala visitation rates were lower than expected for trees with low foliar nitrogen concentrations, but were as expected for trees with high concentrations of nitrogen (Fig. 3).

Although subtle, the differences observed between the foliar chemistry of trees visited by koalas and those available to the animals are the result of reduced rates of koala visitation to the most highly defended trees in the woodland, which also tend to be the largest food patches. Koalas are highly specialized folivores and as such are remarkably tolerant of PSMs. Less-specialized folivores of *Eucalyptus*, which have lower tolerances for FPCs¹⁴, are likely to be

Table 2 | Observed and expected foliar chemistry for all koala visits

	<i>E. globulus</i>		<i>E. viminalis</i>	
	FPC	Nitrogen	FPC	Nitrogen
n_{obs}	907	907	421	421
n_{avail}	484	484	373	373
$\bar{\mu}_{\text{obs}}$	19.18	12.54	16.82	16.16
$\bar{\mu}_{\text{exp}}$	19.72	12.45	17.85	15.97
95% confidence interval	19.41-20.06	12.36-12.53	17.28-18.43	15.80-16.14
$P(\bar{\mu}_{\text{obs}} \geq \bar{\mu}_{\text{exp}})$	<0.0001		<0.0001	
$P(\bar{\mu}_{\text{exp}} \leq \bar{\mu}_{\text{obs}})$	0.017		0.013	

Data has been constrained to the observed pattern of tree size selection. Mean total FPC and nitrogen concentrations ($\bar{\mu}_{\text{obs}}$, in mg g^{-1}) were calculated across all n_{obs} observed koala visits and compared to the expected mean concentrations ($\bar{\mu}_{\text{exp}}$). $\bar{\mu}_{\text{exp}}$ and its 95% confidence interval were estimated by randomly selecting (with replacement) a number of trees from each of four size classes to match the number of trees of each size class visited by koalas.

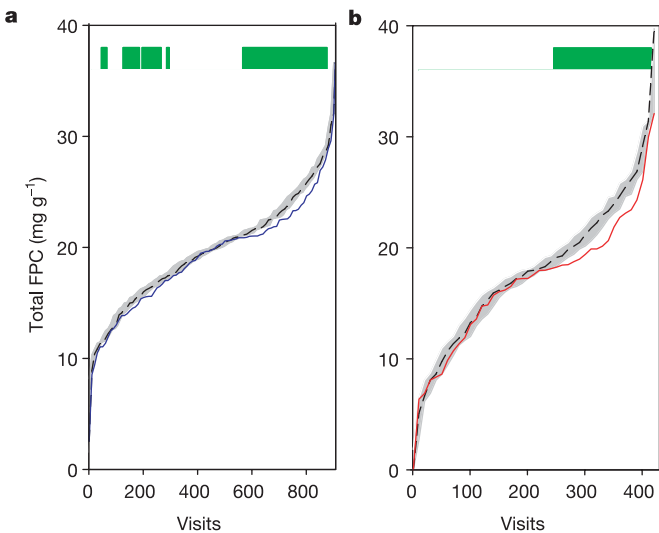


Figure 2 | Distribution of koala visits across FPC concentrations. Observed (coloured lines) and expected (dashed black lines with 95% confidence intervals determined by randomization indicated by grey shading) cumulative distribution functions of tree visitation with respect to FPC concentrations for (a) *E. globulus* and (b) *E. viminalis*. Green bars highlight visits to trees with significantly lower-than-expected FPC concentrations. Data has been constrained to the observed pattern of tree size selection.

deterred from an even larger proportion of trees. The selective herbivory that we have demonstrated may impose evolutionary pressure on the aspects of primary and secondary plant chemistry that we measured. Support for this idea comes from the fact that koala herbivory can seriously affect the fitness of these trees by causing extensive defoliation and mortality²⁵ and that resistance to mammalian browsing and FPC concentrations have a strong genetic basis²⁶. FPC and nitrogen concentrations vary significantly both between^{15,17} and within *Eucalyptus* species, including within-species variation along gradients of altitude and plant productivity²⁷.

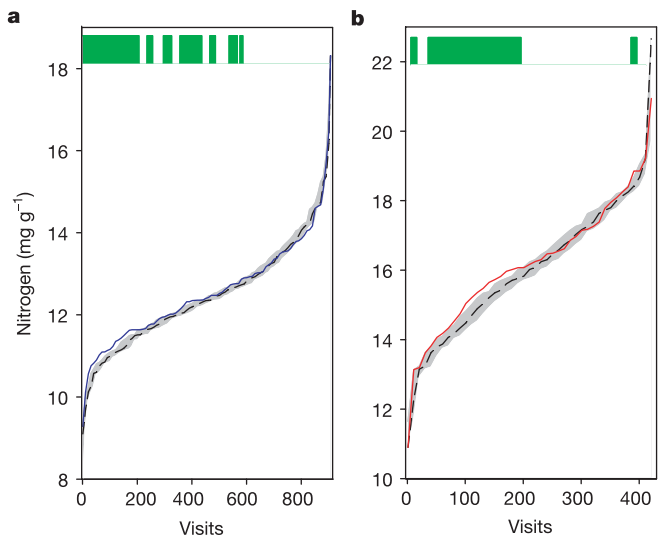


Figure 3 | Distribution of koala visits across nitrogen concentrations. Observed (coloured lines) and expected (dashed black lines with 95% confidence intervals determined by randomization indicated by grey shading) cumulative distribution functions of tree visitation with respect to nitrogen concentrations for (a) *E. globulus* and (b) *E. viminalis*. Green bars highlight visits to trees with significantly greater-than-expected nitrogen concentrations. Data has been constrained to the observed pattern of tree size selection.

Accordingly, we expect that the proportion of trees from which koalas are deterred by high FPC and low nitrogen concentrations varies between sites, providing a mechanism for foliar chemistry to influence the distribution and abundance of this vertebrate herbivore.

METHODS

Koala tree visitation. The study was conducted at the Koala Conservation Centre on Phillip Island (38° 28' S, 145° 13' E). Observations of diurnal tree use by individual koalas were collected at monthly intervals between 1993 and March 2004 by a community group, under the supervision of park rangers. On each occasion, the entire reserve was searched systematically for koalas, and the identities of all koalas found and of the trees occupied were recorded. Approximately 20 koalas were present in the reserve at all times throughout the study. The reserve was enclosed by a koala-proof fence; however, koalas in the reserve forage and breed naturally. Several studies have concluded that tree visitation is a reasonable measure of koala foraging²⁸.

Foliage collection and analysis. All trees in the woodland were individually numbered and mapped, and their circumference at a height of 130 cm recorded. In January 1997, foliage was sampled from the canopy of each *Eucalyptus* tree. Standard methods were used for processing and subsequently for collecting near infrared spectra from these samples^{4,29}. Subsets of samples from each species were analysed for their FPC content and nitrogen concentration using established methods^{4,17}, allowing predictive calibrations to be developed for each of these constituents using Near Infrared Reflectance Spectroscopy (NIRS)²⁹—details of these calibration equations are provided as Supplementary Information.

Randomization tests. Two types of randomization test were used. Initially we tested whether koalas visited trees randomly with respect to size. For each tree species, we generated a list of trees the same length as the list of observed koala visits to that species, by a process of random selection (with replacement) from all trees available. We performed 10,000 iterations of this procedure to determine whether the mean circumference of trees visited by koalas was greater than expected (a one-tailed test), by directly calculating the proportion of randomly generated distributions with means equal to or greater than that observed. Next, we modified the randomization procedure to generate 'expected' cumulative distribution functions of tree visitation with respect to both aspects of foliar chemistry, given the observed pattern of tree visitation with respect to tree size. We divided the trees into four size classes and then randomly selected (with replacement) a number of trees from each class to match the observed number of koala visits to trees of that size class. We performed 10,000 iterations of this procedure in order to generate expected cumulative visitation functions with 95% confidence intervals.

Received 12 January; accepted 15 March 2005.

1. Fraenkel, G. S. The raison d'être of secondary plant substances. *Science* **129**, 1466–1470 (1959).
2. Vour'h, G., Martin, J. L., Duncan, P., Escarre, J. & Clausen, T. P. Defensive adaptations of *Thuja plicata* to ungulate browsing: a comparative study between mainland and island populations. *Oecologia* **126**, 84–93 (2001).
3. Farentinos, R. C., Capretta, P. J., Kepner, R. E. & Littlefield, V. M. Selective herbivory in tassel-eared squirrels: role of monoterpenes in ponderosa pines chosen as feeding trees. *Science* **213**, 1273–1275 (1981).
4. Lawler, I. R., Foley, W. J. & Eschler, B. M. Foliar concentration of a single toxin creates habitat patchiness for a marsupial folivore. *Ecology* **81**, 1327–1338 (2000).
5. Provenza, F. D. Postingestive feedback as an elementary determinant of food preference and intake in ruminants. *J. Range Manage.* **48**, 2–17 (1995).
6. Duncan, A. J., Hartley, S. E. & Iason, G. R. The effect of monoterpene concentrations in Sitka spruce (*Picea sitchensis*) on the browsing behavior of red deer (*Cervus elaphus*). *Can. J. Zool.* **72**, 1715–1720 (1994).
7. Bryant, J. P., Wieland, G. D., Reichardt, P. B., Lewis, V. E. & McCarthy, M. C. Pinosylvin methyl ether deters snowshoe hare feeding on green alder. *Science* **222**, 1023–1025 (1983).
8. Moore, B. D., Foley, W. J., Wallis, I. R., Cowling, A. & Handasyde, K. A.

Eucalyptus foliar chemistry explains koala feeding preferences. *Biol. Lett.* **1**, 64–67 (2005).

9. Provenza, F. D., Villalba, J. J., Dziba, L. E., Atwood, S. B. & Banner, R. E. Linking herbivore experience, varied diets, and plant biochemical diversity. *Small Rum. Res.* **49**, 257–274 (2003).
10. Freeland, W. J. & Janzen, D. H. Strategies in herbivory by mammals: the role of plant secondary compounds. *Am. Nat.* **108**, 269–289 (1974).
11. Schmitz, O. J., Hik, D. S. & Sinclair, A. R. E. Plant chemical defense and twig selection by snowshoe hare: an optimal foraging perspective. *Oikos* **65**, 295–300 (1992).
12. Villalba, J. J., Provenza, F. D. & Bryant, J. P. Consequences of the interaction between nutrients and plant secondary metabolites on herbivore selectivity: benefits or detriments for plants? *Oikos* **97**, 282–292 (2002).
13. Westoby, M. What are the biological bases of varied diets? *Am. Nat.* **112**, 627–631 (1978).
14. Moore, B. D., Wallis, I. R., Marsh, K. J. & Foley, W. J. in *Conservation of Australia's Forest Fauna* (ed. Lunney, D.) 549–575 (Royal Zoological Society of New South Wales, Mosman, NSW, 2004).
15. Eschler, B. M., Pass, D. M., Willis, R. & Foley, W. J. Distribution of foliar formulated phloroglucinol derivatives amongst *Eucalyptus* species. *Biochem. Syst. Ecol.* **28**, 813–824 (2000).
16. Danell, K., Edenius, L. & Lundberg, P. Herbivory and tree stand composition: moose patch use in winter. *Ecology* **72**, 1350–1357 (1991).
17. Moore, B. D. et al. Anti-herbivore chemistry of *Eucalyptus*—cues and deterrents for marsupial folivores. *J. Chem. Ecol.* **30**, 1743–1769 (2004).
18. Hume, I. D. *Marsupial Nutrition* Ch. 5 (Cambridge Univ. Press, Cambridge/New York, 1999).
19. Ullrey, D. E., Robinson, P. T. & Whetter, P. A. Composition of preferred and rejected *Eucalyptus* browse offered to captive koalas, *Phascolarctos cinereus* (Marsupialia). *Aust. J. Zool.* **29**, 839–846 (1981).
20. Degabriele, R. Nitrogen and the koala (*Phascolarctos cinereus*): some indirect evidence. *Aust. J. Ecol.* **8**, 75–76 (1983).
21. Degabriele, R. A relative shortage of nitrogenous food in the ecology of the koala (*Phascolarctos cinereus*). *Aust. J. Ecol.* **6**, 139–141 (1981).
22. White, T. C. R. *The Inadequate Environment: Nitrogen and the Abundance of Animals* Ch. 6.5.1 (Springer, Berlin, 1993).
23. Hindell, M. A., Handasyde, K. A. & Lee, A. K. Tree species selection by free-ranging koala populations in Victoria. *Aust. Wildlife Res.* **12**, 137–144 (1985).
24. Scrivener, N. J. et al. Which trees do wild common brushtail possums (*Trichosurus vulpecula*) prefer? Problems and solutions in scaling laboratory findings to diet selection in the field. *Evol. Ecol. Res.* **6**, 77–87 (2004).
25. Martin, R. W. Overbrowsing, and decline of a population of the koala, *Phascolarctos cinereus*, in Victoria. I. Food preference and food tree defoliation. *Aust. Wildlife Res.* **12**, 355–365 (1985).
26. O'Reilly-Wapstra, J. M., McArthur, C. & Potts, B. M. Genetic variation in resistance of *Eucalyptus globulus* to marsupial browsers. *Oecologia* **130**, 289–296 (2002).
27. Moore, B. D., Wallis, I. R., Wood, J. & Foley, W. J. Foliar nutrition, site quality and temperature affect foliar chemistry of tallowood (*Eucalyptus microcorys*). *Ecol. Monogr.* **74**, 553–568 (2004).
28. Martin, R. W. & Handasyde, K. A. The koala: natural history, conservation and management. Ch. 4 (Univ. New South Wales Press, Sydney, 1999).
29. McIlwee, A. M., Lawler, I. R., Cork, S. J. & Foley, W. J. Coping with chemical complexity in mammal-plant interactions: near-infrared spectroscopy as a predictor of *Eucalyptus* foliar nutrients and of the feeding rates of folivorous marsupials. *Oecologia* **128**, 539–548 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to A. Reed and 'Friends of the Koalas Inc.' for the use of their koala data; I. Lawler, K. Handasyde, I. Wallis, R. Bednarik, M. Ebberts and K. Marsh for assistance in the field and laboratory; Phillip Island Nature Park for allowing access to the Koala Conservation Centre and for providing accommodation to B.D.M.; the Australian Research Council for funding to W.J.F.; and C. Johnson, M. Jennions, J. DeGabriel and M. Symonds for comments on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.D.M. (Ben.Moore@jcu.edu.au).

Cattle movements and bovine tuberculosis in Great Britain

M. Gilbert¹, A. Mitchell², D. Bourn³, J. Mawdsley², R. Clifton-Hadley² & W. Wint³

For 20 years, bovine tuberculosis (BTB) has been spreading in Great Britain (England, Wales and Scotland) and is now endemic in the southwest and parts of central England and in southwest Wales, and occurs sporadically elsewhere. Although its transmission pathways remain poorly understood, the disease's distribution was previously modelled statistically by using environmental variables and measures of their seasonality¹. Movements of infected animals have long been considered a critical factor in the spread of livestock diseases, as reflected in strict import/export regulations, the extensive movement restrictions imposed during the 2001 foot-and-mouth disease outbreak^{2,3}, the tracing procedures after a new case of BTB has been confirmed and the Government's recently published strategic framework for the sustainable control on BTB⁴. Since January 2001 it has been mandatory for stock-keepers in Great Britain to notify the British Cattle Movement Service of all cattle births, movements and deaths⁵. Here we show that movements as recorded in the Cattle Tracing System data archive, and particularly those from areas where BTB is reported, consistently outperform environmental, topographic and other anthropogenic variables as the main predictor of disease occurrence. Simulation distribution models for 2002 and 2003, incorporating all predictor categories, are presented and used to project distributions for 2004 and 2005.

BTB, once almost eradicated from Great Britain, has been spreading since 1984, and the number of detected BTB cases continues to rise exponentially (Fig. 1). Because BTB transmission cycles remain poorly understood, the causes of this epidemic are keenly debated, with various possible explanations including transmission from wildlife reservoirs, inadequate control measures, agro-environmental factors and movement of infected animals^{6–8}. Previous work has shown the value of environmental variables and measures of their seasonality⁹, as encapsulated in Fourier-processed satellite imagery¹⁰, in modelling the distributions of BTB in Great Britain¹. The pattern of spread of BTB between 1984 and 2003 shows an expanding core area with outlying foci (Fig. 1). Such a pattern is commonly described in invasion ecology by stratified dispersal models that combine short-distance spread with long-distance dispersal events^{11–13}. For infectious diseases of livestock, short-distance spread can be viewed as contagion to adjacent or nearby farms located within a few kilometres, by direct contact or borne by wind, insects, rodents or alternative hosts, and generally resulting in the local spatial clustering of cases. Long-distance jump-spread can be viewed as contagion occurring between locations separated by large areas of disease absence, and caused by movements of infectious individuals or infected material. The present study sought to evaluate the relative importance of cattle movement as a predictor of BTB distribution and invasion in Great Britain, and to use the established models for short-term predictions.

Two sets of analyses were undertaken. The first was designed to

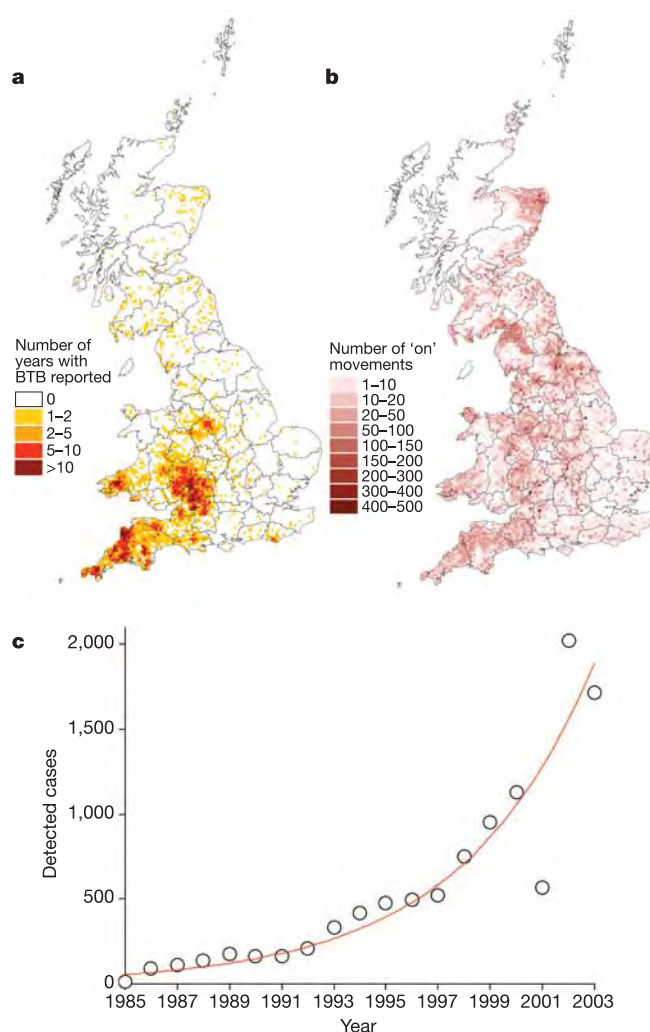


Figure 1 | Distribution and spread of BTB in 1985–2003 and distribution of annual cattle movements. **a**, Distribution of the number of years in which BTB has been reported. **b**, Distribution of cattle inward movements in 2002 by 5-km cells. **c**, Change in the number of BTB detected cases. Note that the number of cases recorded for 2001 was relatively low, probably because of the reduction in testing during the outbreak of foot and mouth disease. They seem to have been compensated for by a comparative increase in 2002. The equations of the curve are: $y = 8.70 \times 10^{-168} e^{0.1958x}$; $R^2 = 0.876$; $n = 19$; $P < 0.001$.

¹Biological Control and Spatial Ecology CP160/12, Université Libre de Bruxelles, avenue F.D. Roosevelt 50, B-1050 Brussels, Belgium. ²Veterinary Laboratories Agency, Weybridge, New Haw, Addlestone, Surrey KT15 3NB, UK. ³Environmental Research Group Oxford Limited, PO Box 346, Oxford OX1 3QE, UK.

Table 1 | Summary statistics and top five predictors for models of BTB during 2002 and 2003

Parameter	Year			
	2002	2003	2002	2003
Movement from	2002	2003	2000	2001
Variables available to model	100	100	100	100
Variables included in model	18	21	20	20
−2 log-likelihood	2,671.8	3,637.5	3,000.3	3,657.3
χ^2	4,444.4	2,730.4	4,174.2	2,748.7
P	<0.001	<0.001	<0.001	<0.001
Correct presence (%)	90.04	87.07	91.28	85.51
Correct absence (%)	91.82	81.67	91.07	83.15
Overall correct (%)	90.94	84.39	91.17	84.33
Overall kappa	0.82	0.69	0.82	0.69
Variable 1	PMI (1,033.4)	PMI (990.4)	PMI (842.1)	PMI (970.0)
Variable 2	DBTD00 (48.4)	ATVM (74.0)	DBTD00 (61.5)	ATVM (73.7)
Variable 3	ATVARA (47.2)	ATMN (44.6)	NDVIPH1 (34.6)	ATMN (44.9)
Variable 4	NDVIMAX (29.1)	ATPH1 (30.5)	DBR (34.0)	NDVIMN (33.6)
Variable 5	ATRNG (16.9)	MIRMN (28.2)	NDVIMAX (29.1)	VPDPH2 (25.9)

PMI, proportion of movements from infected areas. DBTD00, distance to the nearest cell with BTB present in year 2000. ATVARA, air temperature (variance all). NDVIMAX, normalized deviation vegetation index (maximum). ATRNG, air temperature (range). ATVM, air temperature (variance of mean). ATMN, air temperature (mean). ATPH1, air temperature (phase 1). MIRMN, middle infrared reflectance (mean). NDVIPH1, normalized deviation vegetation index (phase 1). DBR, distance to nearest recorded badger presence. NDVIMN, normalized deviation vegetation index (mean). VPDPH2, vapour pressure deficit (phase 2). Figures in parentheses are variable Wald statistics. The model with all predictors is provided in Supplementary Table 2.

extend the earlier modelling of known BTB presence using multiple logistic regressions, to compare cattle movement indices with other previously assessed environmental, demographic, agricultural and climatic parameters (detailed in Supplementary Information). Procedures were developed¹⁴ to extract additional parameters from the 97-million-record Cattle Tracing System (CTS) data archive⁵. These cattle movement data for 2000–03 inclusive were incorporated into the suite of predictor variables available to model BTB distributions. With disease data for 2002 and 2003, the analyses showed unequivocally that movement parameters consistently outperform the other variables in predicting BTB distributions (as quantified by their Wald statistic compared with other predictors), and can produce model accuracies with kappa values of about 0.7 and above¹⁵ (Table 1). Of the range of movement variables tested (total number of 'on' movements, number of movements from infected areas and the proportion of movements from infected areas) the proportion of movements from infected areas arriving at a location was found to be most closely associated with disease presence. The predictive power of this variable also exceeds that of distance to previous disease cases, used as an indicator of previous disease status in the vicinity. The analyses were repeated with movement variables from two years earlier (Table 1). These historical variables, which also provided accurate distribution models, were then replaced with concurrent values, and the models were rerun to generate short-term BTB predictions for 2004 and 2005. The modelled distributions and corresponding projections are shown in Fig. 2, together with insets

of actual distributions. High kappa values indicate a good match between actual and modelled predictions. The model predictions indicate that, although below the usually accepted 50% threshold for presence, some areas in northeast Wales, Cumbria and the Scottish borders have a 20–30% risk level of BTB occurrence. These areas stand out more in the projections because the patches are larger and have somewhat higher projected risks, implying that, should current trends continue, the disease is more likely to be found there.

However, the projections are heavily dependent on disease distribution in the base year, and any abnormality is reflected in projections derived from it. This can only be avoided by developing process-based models—for which insufficient epidemiological detail is currently available—or by developing projection models based on more than one year's disease data.

Accordingly, a second set of analyses focused on assessing whether stochastic simulation methods of estimating disease spread could provide a dynamic disease spread model. This model, when applied to known starting BTB distributions, would replicate the observed disease spread and, given the use of predictors derived from a time series rather than a single year, could be projected into the medium-term future. Because the spatial and temporal patterns of movement were shown to be consistent from year to year, it was possible to assemble a model incorporating disease data from 1997 to 2003, with a set of predictor variables. To address the probably stratified nature of the disease's dispersal, multi-annual logistic regressions of BTB occurrence were produced for core and remote areas (core areas

Table 2 | Multi-annual multiple logistic regression of BTB occurrence in Great Britain from 1990 to 2003

Parameter	Core	Remote with movement data	Remote with transformed distance
Variables available to model	100	100	100
Variables included in model	2	6	6
−2 log-likelihood	10,612.3	7,196.0	6,255.1
χ^2	1,316.7	5,552.5	6,493.5
P	<0.001	<0.001	<0.001
Correct presence (%)	56.31	79.07	87.67
Correct absence (%)	78.03	86.39	84.51
Overall correct (%)	67.15	82.83	86.04
Overall kappa	0.34	0.66	0.72
Variable 1	NYBTB (689.1)	NYBTB (129.2)	NYBTB (80.7)
Variable 2	DNT (268.6)	DNT (313.5)	DNT (184.7)
Variable 3	–	PMI (451.8)	TDBTD (1,150.1)
Variable 4	–	NDVIMN (87.9)	NDVIMN (54.0)
Variable 5	–	CAD (145.9)	CAD (36.2)
Variable 6	–	PCU (135.0)	PCU (73.2)

NYBTB, number of years of past BTB infection in the 5-km cell. DNT, number of infected 1-km cells in the previous year in a doughnut window 5 km in radius. CAD, cattle density. PCU, percentage of cultivation and managed grassland. TDBTD, transformed distance to the nearest 1 km square with BTB reported in the previous year. Figures in parentheses are Wald statistics.



Parameter estimate period	Overall correct (%)	Correct presence (%)	Correct absence (%)	New cases correct (%)	-2 log-likelihood	R ²	Threshold
PMI model							
1997-2002	84.6	83.3	84.7	81.6	-8,961	0.251	0.010
1997-2001	83.4	82.3	83.4	80.6	-9,408	0.214	0.003
1997-2000	83.9	84.2	83.9	82.6	-9,048	0.244	0.009
TDBTB model							
1997-2002	83.2	82.4	83.2	81.0	-9,165	0.234	0.010
1997-2001	83.5	81.7	83.5	80.4	-9,311	0.222	0.006
1997-2000	84.2	81.8	84.2	80.3	-9,110	0.239	0.010

493

being defined as those where BTB had been detected in two or more of the three previous years; see Table 2 and Supplementary Information).

BTB occurrence in core areas was associated with the previous year's BTB status and with disease status in the previous year in surrounding areas (Table 2). In remote areas, BTB occurrence was associated with the following: variables describing past BTB status; the status of the disease in the previous year in surrounding areas; the proportion of inward movements from infected areas; the density of cattle; the proportion of grassland; and the mean of the normalized deviation vegetation index. Local persistence, short-distance spread from adjacent locations, inward movements from infected areas, host density and two other variables relating to land use and vegetation seem to be significant factors associated with the establishment of new infections away from established foci. The predictions in remote areas are better than in core areas (as quantified by their higher kappa value; Table 2) probably because the latter represent regions of endemicity, where changes in disease status are more difficult to predict consistently by using environmental variables. The possible introduction of BTB by animal movements might also be less critical in such areas than in those where the disease has yet to become established.

The distribution modelling demonstrates the value of movement indicators in short-term projections, and the simulation analyses validate medium-term projections and suggest that the multi-annual

models can reproduce the known spread of BTB effectively (Table 3, Fig. 3). Models using parameters derived from an analysis of BTB distribution in the periods 1997–2002 and 1997–2000 provided reasonably good predictions, whereas the model using the period 1997–2001 for determining the parameters tended to underestimate the distribution of BTB in 2003. This underestimation results from the smaller number of detected cases in 2001 (Fig. 1c), which reduces the probability of BTB presence in the modelled time-series. The model including the proportion of movement from infected areas (PMI) best reproduces the observed distribution in 2003, followed by the model in which the distance from nearest infected area is weighted by the frequency distribution of movement distances (TDBTB), which tends to underestimate the observed spread (Table 3, Fig. 3b, d). Although the TDBTB model could be used to generate predictions (Fig. 3e, f), a comparison of the outputs of the TDBTB and PMI models indicates that better predictions would be generated by a PMI model for which the proportion of inward movement from infected areas was estimated at each simulation run rather than extracted from a fixed distribution. This would require 'on-the-fly' extraction of inward movements from the CTS database, which, although computationally intensive, might justify further exploration. With a present maximum of only four predictors, other than variables concerning the past status of BTB, the predictive power of the simulations might also be improved by incorporating additional variables; the limitation is computational processing time.

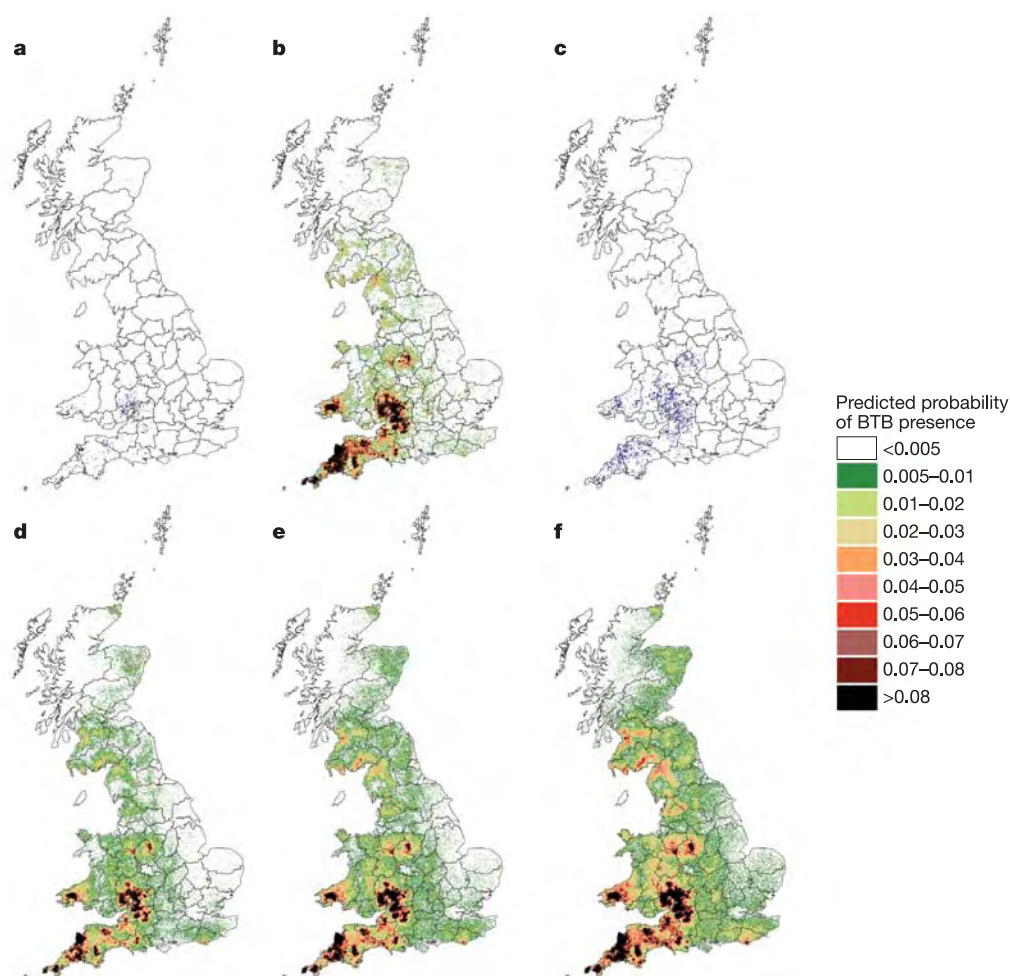


Figure 3 | Observed and predicted distributions of BTB. **a**, Starting distribution in 1997. **b**, Distribution predicted by the PMI model for 2003, using the 1997–2002 periods for estimates of parameters. **c**, Observed 2003 distribution. **d–f**, Predicted distribution for 2003 (**d**), 2004 (**e**) and 2005 (**f**)

using the TDBTB model and the 1997–2002 periods for estimates of parameters. The average numbers of predicted cases by the TDBTB model were 1,750, 2,224 and 2,799 for 2003, 2004 and 2005, respectively.

The British Cattle Movement Service and the CTS were set up to ensure the identification and traceability of individual cattle during the recovery period after the BSE crisis; they were not intended to serve as a disease-control support system for fast-moving diseases, such as the outbreak of foot-and-mouth disease that occurred in 2001 (ref. 2). However, as the number of detected BTB cases continues to rise exponentially (Fig. 1c), the need to identify critical risk factors becomes ever more important. These results demonstrate that the movement of animals, especially from locations where BTB is present and particularly to locations outside endemic core areas, is one such critical factor.

These findings support the case for movement controls, especially from 'core' to 'remote' locations, as a disease control measure. The models developed here can be used to assess the potential impact of various movement control strategies and so provide underpinning for cost-benefit analyses of this approach to controlling BTB.

However, it should be pointed out that despite the demonstrated association of movement levels with the disease, there are several regions into which many animals are imported where the disease appears regularly but where it does not seem to persist. Potential explanations include the possibility that the critical movement categories (if any) do not occur in these areas, the fact that the imported animals may remain in them only briefly before being slaughtered or moved elsewhere, and that either suitable wildlife reservoirs do not exist in those areas or that conditions necessary for establishment in those species are not fulfilled.

This work has therefore established a clear requirement for further examination of cattle movements to define critical movement categories and to incorporate length of stay into the predictive models, both of which are readily feasible using the methods presented here. Combining cattle movement data, molecular typing data of *Mycobacterium bovis* isolates and knowledge of local wildlife BTB infection status would provide a sound basis for investigating why disease becomes more widespread or persistent in some areas and not others.

METHODS

Data. BTB data were derived from the UK Department for Environment, Food and Rural Affairs (DEFRA) VETNET database for the period 1984–2003 and included BTB monitoring data for more than 90,000 holdings in Great Britain. Two sources of uncertainty are acknowledged in the BTB monitoring data. First, imperfect sensitivity of BTB skin tests may result in a small fraction of false negatives. Positive skin tests in an animal are confirmed only if *M. bovis* typical lesions are identified, or if a culture positive for *M. bovis* is obtained. For a herd breakdown to be considered confirmed it must contain one or more confirmed animals. The proportion of confirmed herd breakdowns was 65.5%, 71.8%, 61.9% and 55.7% in the years 2000, 2001, 2002 and 2003, respectively. Second, BTB testing frequency is not homogeneously distributed in Great Britain: the testing frequency is once every one to four years, which may result in uncertainty on the dating of detected cases in low-frequency testing areas.

Predictors included a broad range of anthropogenic, biological, demographic, climatic and topographic variables compiled and resampled at 1 km resolution (Supplementary Table 1). Variables representing disease persistence were added to the predictor suite, including the number of years of past BTB infection in a 5-km cell (a measure of temporal local persistence) and the number of infected cells in the previous year within a 5-km doughnut window around the sampling point (as a measure of short-distance spread).

Preliminary treatments of the raw CTS data included the geo-referencing of movement records (98% of locations associated with cattle movement were geo-referenced), pairing movement records¹⁴ (all events except birth and death have two records: one for the location from which the animal had moved and the other for the location onto which the animal had moved; for any analysis of movement patterns, 'off' and 'on' records have to be 'paired'). Movement data from the period 2000–03 were mapped and added to the series of predictors obtained for each 1-km and 5-km cell. These variables included the total number of movements into a cell, the total number of movements from an infected area, and the proportion of movements that originated from infected areas.

Analysis. The association between BTB occurrence and the predictors was explored using a stepwise multiple logistic regression analysis of 2002 and 2003 BTB distribution data. Data values were extracted for data points

corresponding to all BTB-positive locations and an equal number of BTB-negative locations systematically (that is, every n th negative value within a geographically sorted data set) covering the whole land mass. Models were assessed using the kappa index of agreement¹⁵.

A simulation model was also developed that, when applied to a known starting distribution, would replicate the known spread of the disease and could be projected into the future. Simulation modelling has been used previously to identify landscape characteristics associated with the spread of diseases¹⁶ or invading organisms¹³ and to forecast their invasion pattern¹⁷. The simulation model was built in two main steps.

First, it was necessary to identify the predictors applying to the disease over several years. This involved building a multi-annual database of disease presence and absence, and defining the best predictors and their numeric multipliers. Cattle movement data were available only for the period 2000–03, but the spatial distribution of cattle movement as measured by the number of annual inward movements per 5-km pixel was relatively stable in this period (see Supplementary Information). It was therefore assumed that movement patterns were similar in the previous years, allowing a fixed movement index—the proportion of movement between 2000 and 2003 from infected areas—to be incorporated into the multi-annual predictor suite. In an attempt to differentiate the behaviour of the disease in its focal areas from that elsewhere, and as a first step towards pinpointing where new foci might become established, predictors were identified separately for 'core' and 'remote' areas, with the 'core' areas defined as those cells in which the disease had been found in at least two of the previous three years. Two separate multi-annual multiple logistic regression models were thus built for core and remote areas, respectively, using the observed distribution of BTB in the period 1990–2003 as a dependent variable. Variables relating to local persistence (number of years of past BTB infection in the 5-km cell), to short-distance spread (number of infected cells in the previous year in a doughnut-shaped window 5 km in radius), to animal movements (total movements in, total movements in from infected areas, proportion of movements from infected areas), and to cattle density were entered first; other variables were then entered into the model with a standard forward-entry stepwise procedure. We sought to restrict the model to variables with the highest predictive power, and only those presenting more than 1% of log-likelihood change after removal were retained. Finally, we ensured that all variables used were consistently associated with the presence or absence of BTB over time, in other words that were significant within each year when tested on annual models, by rejecting variables found not significant for more than one year in the period 1997–2003.

The second step involved using the selected variables to simulate the spread of BTB from a starting year. The parameter of each variable was established from multi-annual multiple logistic regression models detailed above, with the incorporation of the year as predictor to account for the trend in increased case numbers. The algorithm used to simulate the spread of BTB involved the following: first, the infection probability of each cell was estimated as a logistic function of the predictors identified through the multi-annual logistic regressions of disease data from 1997 to 2003; second, a layer of random numbers uniform over the interval [0–1] was generated and cells with a random number lower than their infection probability were set as BTB-present; and third, each cell's BTB status was updated and the algorithm reiterated to simulate the spread in the next year. Processing constraints and the current structure of the CTS database prevented movement data to be generated on the fly during the simulation process as a function of the previous year's simulated distribution; it was therefore necessary to identify a surrogate variable for animal movement from previously infected areas. The animal movement kernel, averaged over the years 2000–03, was modelled as a function of distance (see Supplementary Information). The function obtained was used at simulation time to transform geographical distance to the nearest square infected in the previous year into a surrogate index of potential inward movements. The algorithm started with the observed distribution of BTB in a first year and was iterated until the target year. This set of n hypothetical distributions of BTB in n consecutive years constitutes a single run, of which 500 were performed; these distributions were averaged over the 500 runs to derive the predicted probability of presence at 1 km resolution. This was compared to the observed distribution at the same scale by estimating the model -2 log-likelihood and McFadden's pseudo- R^2 ; the threshold used to compare predicted and observed presence and absence was determined so as to minimize the difference between the percentage of correct presence and the percentage of correct absence. The model was tested for its ability to predict the distribution in 2003, using parameters derived from the multi-annual logistic regression analysis of the periods 1997–2002, 1997–2001 and 1997–2000, so as to test the model's ability to predict the BTB distribution in the last year of the training set (2003), and then for 1–3 years ahead.

Received 20 January; accepted 18 March 2005.

- Wint, G. R. *et al.* Mapping bovine tuberculosis in Great Britain using environmental data. *Trends Microbiol.* **10**, 441–444 (2002).
- Gibbens, J. C. *et al.* Descriptive epidemiology of the 2001 foot and mouth disease epidemic in Great Britain: The first five months. *Vet. Rec.* **149**, 729–743 (2001).
- Haydon, D. T. *et al.* The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth disease outbreak. *Proc. R. Soc. Lond. B* **270**, 121–127 (2003).
- DEFRA. *Government Strategic Framework for the Sustainable Control of Bovine Tuberculosis (bTB) in Great Britain* March (Department for Environment, Food and Rural Affairs, London, 2005).
- NAO. *Identifying and Tracking Livestock in England* (Stationery Office, on behalf of National Audit Office, London, 2003).
- Bourne, F. J. *et al.* *An Epidemiological Investigation into Bovine Tuberculosis—Towards a Sustainable Policy to Control TB in Cattle* (Department for Environment, Food and Rural Affairs, London, 2001).
- Krebs, J. *et al.* *Bovine Tuberculosis in Cattle and Badgers* (Ministry of Agriculture, Fisheries and Food Publications, London, 1997).
- Donnelly, C. A. *et al.* Impact of localized badger culling on tuberculosis incidence in British cattle. *Nature* **426**, 834–837 (2003).
- Hay, S. I. & Lennon, J. J. Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Trop. Med. Int. Health* **4**, 58–71 (1999).
- Hay, S. I., Randolph, S. E. & Rogers, D. J. (eds) *Remote Sensing and Geographical Information Systems in Epidemiology (Advances in Parasitology Vol. 47)* (Academic, London, 2000).
- Hengeveld, R. *Dynamics of Biological Invasions* (Chapman & Hall, London, 1989).
- Shigesada, N. & Kawasaki, K. *Biological Invasions: Theory and Practices* (Oxford Univ. Press, Oxford, 1997).
- Gilbert, M., Grégoire, J.-C., Freise, J. & Heitland, W. Long-distance dispersal and human population density allow the prediction of invasive patterns in the horse-chestnut leafminer *Cameraria ohridella*. *J. Anim. Ecol.* **73**, 459–468 (2004).
- Mitchell, A. *et al.* Characteristics of cattle movements in Britain—an analysis of records from the Cattle Tracing Scheme. *J. Anim. Sci.* **80**, 265–273 (2005).
- Congalton, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **37**, 35–46 (1991).
- Smith, D. L., Lucey, B., Waller, L. A., Childs, J. E. & Real, L. A. Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc. Natl. Acad. Sci. USA* **99**, 3668–3672 (2002).
- Russell, C. A., Smith, D. L., Waller, L. A., Childs, J. E. & Real, L. A. A priori prediction of disease invasion dynamics in a novel environment. *Proc. R. Soc. Lond. B* **271**, 21–25 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Cox for comments on an early draft of the paper; D. Rogers and his TALA Research Group at Oxford for advice and for providing satellite imagery; and P. Roeder for help and encouragement. This work was funded by a DEFRA research project.

Author Contributions M.G. did the simulation modelling and spatial statistics. W.W. did the data management, analysis and regression modelling. A.M. and J.M. did the movement data acquisition, processing and analysis. D.B. and R.C.-H. did the interpretation, editing and project management. A.M., J.M. and R.C.-H. are Civil Servants and as such their work is subject to Crown Copyright.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.W. (william.wint@zoo.ox.ac.uk).

Allosteric modulation of the presynaptic Ca^{2+} sensor for vesicle fusion

Xuelin Lou¹, Volker Scheuss^{1†} & Ralf Schneggenburger¹

Neurotransmitter release is triggered by an increase in the cytosolic Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$), but it is unknown whether the Ca^{2+} -sensitivity of vesicle fusion is modulated during synaptic plasticity. We investigated whether the potentiation of neurotransmitter release by phorbol esters^{1–3}, which target presynaptic protein kinase C (PKC)/munc-13 signalling cascades^{4–6}, exerts a direct effect on the Ca^{2+} -sensitivity of vesicle fusion. Using direct presynaptic Ca^{2+} -manipulation and Ca^{2+} uncaging at a giant presynaptic terminal, the calyx of Held, we show that phorbol esters potentiate transmitter release by increasing the apparent Ca^{2+} -sensitivity of vesicle fusion. Phorbol esters potentiate Ca^{2+} -evoked release as well as the spontaneous release rate. We explain both effects by an increased fusion ‘willingness’ in a new allosteric model of Ca^{2+} -activation of vesicle fusion. In agreement with an allosteric mechanism, we observe that the classically high Ca^{2+} cooperativity in triggering vesicle fusion (~ 4) is gradually reduced below $3 \mu\text{M}$ $[\text{Ca}^{2+}]_i$, reaching a value of <1 at basal $[\text{Ca}^{2+}]_i$. Our data indicate that spontaneous transmitter release close to resting $[\text{Ca}^{2+}]_i$ is a consequence of an intrinsic property of the molecular machinery^{7,8} that mediates synaptic vesicle fusion.

We studied transmitter release modulation at the calyx of Held, a large excitatory synapse of the central nervous system (CNS) that is uniquely accessible to manipulations of the presynaptic intracellular Ca^{2+} concentration^{9–11}. Bath application of the phorbol ester phorbol-12,13-dibutyrate (PDBu; $1 \mu\text{M}$) led to a potentiation of evoked excitatory postsynaptic currents (EPSCs), which reached a near-maximal value of $227 \pm 14\%$ (mean $\pm$ s.e.m.) of control after 200 s (Fig. 1a, b). PDBu also increased the frequency of spontaneous, miniature EPSCs (mEPSCs; Fig. 1c), from 1.27 ± 0.3 Hz for control to 5.1 ± 2.1 Hz following bath-application of PDBu. The mean mEPSC amplitudes were unchanged (Fig. 1c, d), indicating that PDBu potentiates synaptic transmission by means of a presynaptic mechanism⁴.

To investigate directly whether PDBu potentiates transmitter release by increasing the Ca^{2+} -sensitivity of vesicle fusion^{3,12}, we made simultaneous pre- and postsynaptic whole-cell recordings, and evoked transmitter release by presynaptic Ca^{2+} uncaging (Fig. 2). Under these conditions of presynaptic whole-cell recordings, the potentiation of EPSCs by PDBu was transient (see Supplementary Fig. 1), an effect that we attribute to the loss of presynaptic signalling molecules and/or the loss of glutamate from the presynaptic terminal¹³. In the experiments shown below, we therefore restricted the analysis to times around the peak of the PDBu-induced potentiation.

EPSCs evoked by weak Ca^{2+} uncaging stimuli, which elevated the presynaptic $[\text{Ca}^{2+}]_i$ to 2.5 – $5 \mu\text{M}$, were strongly potentiated by $1 \mu\text{M}$ PDBu (Fig. 2a). We deconvolved the EPSCs with the estimated mEPSC waveform to derive the transmitter release rates (Fig. 2a; see Methods). The release rate traces in the example shown in Fig. 2a (bottom panel) indicate that transmitter release was potentiated

2.6-fold. On average, we observed a 4.4 ± 1.04 -fold potentiation of peak transmitter release rates for flashes that elevated $[\text{Ca}^{2+}]_i$ to 2.5 – $5 \mu\text{M}$ ($n = 6$ cells). The integrated transmitter release rate traces (Fig. 2b) could often be separated into a fast and a slow component of release, on the basis of double-exponential fits. In those cells in which the separation was feasible at low $[\text{Ca}^{2+}]_i$ (2.5 – $5 \mu\text{M}$; $n = 3$ out of 6 cells), the amplitude of both the fast and the slow component of release was increased to $288 \pm 57\%$ and $432 \pm 117\%$ of control, respectively.

When transmitter release was evoked by stronger Ca^{2+} -uncaging stimuli that elevated presynaptic $[\text{Ca}^{2+}]_i$ to 10 – $13 \mu\text{M}$, we found that PDBu did not notably potentiate the EPSC amplitudes (Fig. 2c). With these stronger $[\text{Ca}^{2+}]_i$ stimuli, the EPSC rose rapidly and had large amplitudes, indicative of the rapid release of a large number of readily releasable vesicles. Application of $1 \mu\text{M}$ PDBu did not lead to a significant increase in EPSC amplitude or peak transmitter release rates under these conditions (Fig. 2c). However, the rise times of

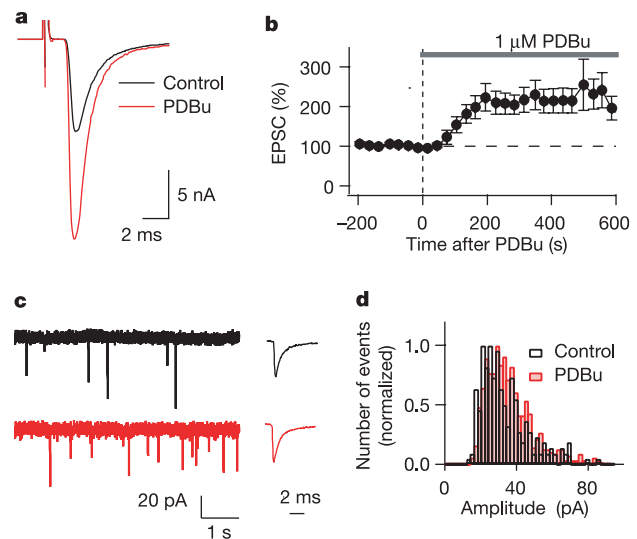


Figure 1 | Potentiation of evoked EPSCs and miniature EPSC frequency by the phorbol ester PDBu. **a**, Evoked EPSCs before (black trace), and after (red trace) bath application of $1 \mu\text{M}$ PDBu. **b**, Average time course of EPSC potentiation by PDBu (mean $\pm$ s.e.m.; $n = 4$ cells). **c**, Postsynaptic current before and after application of PDBu. Note that the frequency of mEPSCs increased, but mEPSC amplitude (average traces shown on the right) was virtually unchanged. **d**, mEPSC amplitude histogram before (black) and after (red) application of PDBu. The average mEPSC amplitude was 36.1 pA for control conditions and 38.4 pA after PDBu application. Data shown in **a**, **c** and **d** are from the same cell.

¹AG Synaptische Dynamik und Modulation und Abt. Membranbiophysik, Max-Planck-Institut für biophysikalische Chemie, Am Fassberg 11, D-37077 Göttingen, Germany.

[†]Present address: Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

EPSCs were shortened by PDBu application (Fig. 2c, inset), indicating that PDBu slightly speeds up the release process. As $[Ca^{2+}]_i$ steps to $10 \mu M$ or more rapidly deplete a pool of readily releasable vesicles¹⁰, the small effect of PDBu on EPSCs evoked by strong Ca^{2+} -uncaging stimuli indicates that PDBu does not induce a marked increase in the pool size. With pool-depleting $[Ca^{2+}]_i$ steps

to $10\text{--}13 \mu M$, neither the fast nor the slow component of transmitter release was significantly increased by PDBu (Fig. 2d; $102 \pm 5\%$ and $105 \pm 10\%$ of control, respectively; $n = 7$ pairs each; $P > 0.5$). Similar results were obtained by analysing the cumulative transmitter release rates evoked by pool-depleting presynaptic depolarizations (see Supplementary Fig. 2). These findings indicate that in the calyx of Held, unlike in chromaffin cells¹⁴ and hippocampal neurons¹⁵, PDBu does not significantly increase the size of a readily releasable vesicle pool. Furthermore, PDBu does not change the relative weighting of a rapidly releasable, and a more slowly releasable sub-pool of readily releasable vesicles¹⁶.

Figure 2e summarizes the experiments shown in Fig. 2a–d by plotting the transmitter release rates as a function of $[Ca^{2+}]_i$, both for the control conditions and after applying $1 \mu M$ PDBu. The control data (Fig. 2e, black circles, $n = 23$ cell pairs) has a slope of 4.2 ± 0.5 (slope $\pm 95\%$ confidence interval), as revealed by linear regression after logarithmic transformation of the data set. Conversely, the data obtained in the presence of PDBu (open triangles) has a slope of 3.3 ± 0.3 , indicating that PDBu reduces the effective cooperativity with which Ca^{2+} induces vesicle fusion. Figure 2f shows the effect of PDBu relative to the line fitted to the control data in each cell (see Methods). For small $[Ca^{2+}]_i$ steps (up to $1\text{--}3 \mu M$), PDBu causes a 4–5-fold increase in release rate ($P < 0.001$; one-sample t -test). The

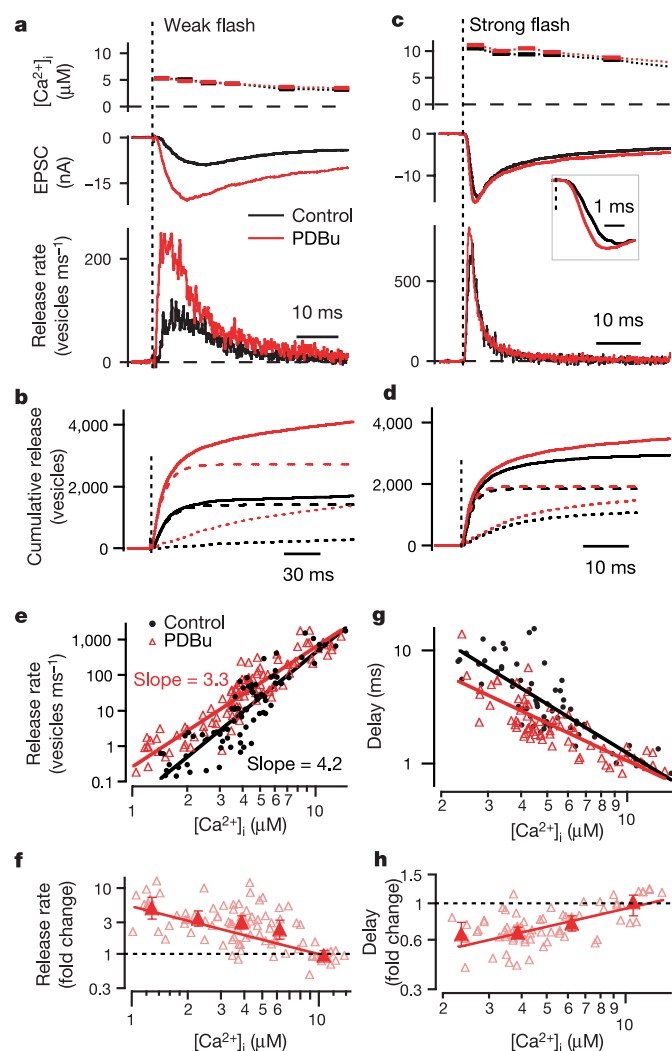


Figure 2 | Presynaptic Ca^{2+} uncaging shows that phorbol ester increases the Ca^{2+} sensitivity of vesicle fusion, concomitant with a reduced Ca^{2+} cooperativity of vesicle fusion. **a**, Responses to a weak Ca^{2+} -uncaging stimulus. The upper, middle and lower panels show presynaptic $[Ca^{2+}]_i$, EPSCs and transmitter release rate, respectively. **b**, Integrated transmitter release rate traces for the responses in **a**. The fast (dashed lines) and slow (dotted lines) components of release are shown. **c**, Same as **a**, but for a strong Ca^{2+} -uncaging stimulus in a different cell pair. The inset shows the EPSC on an expanded time scale. **d**, Same as **b**, but for the responses shown in **c**. In **a–d**, control responses are shown in black and responses around the peak of the PDBu-potentiation are shown in red. **e**, Double-logarithmic plot of peak transmitter release rates (as measured in **a** and **c**) as a function of presynaptic $[Ca^{2+}]_i$. Lines in **e–h** are linear regression fits to the logarithmically transformed data sets ($n = 23$ cell pairs). **f**, Potentiation of peak release rates by PDBu, relative to the fitted line of the control data points in each cell (see Methods). Here (and in **h**) the individual data points (open symbols) were grouped and averaged (filled symbols), with error bars representing the 95% confidence interval. All groups were significantly different from 1 ($P < 0.003$; one-sample t -test), except for the data at the highest $[Ca^{2+}]_i$ ($P \sim 0.5$) in each plot. **g**, Synaptic delays as a function of $[Ca^{2+}]_i$, both under control conditions (black) and in the presence of $1 \mu M$ PDBu (red). **h**, Relative change in synaptic delays induced by $1 \mu M$ PDBu.

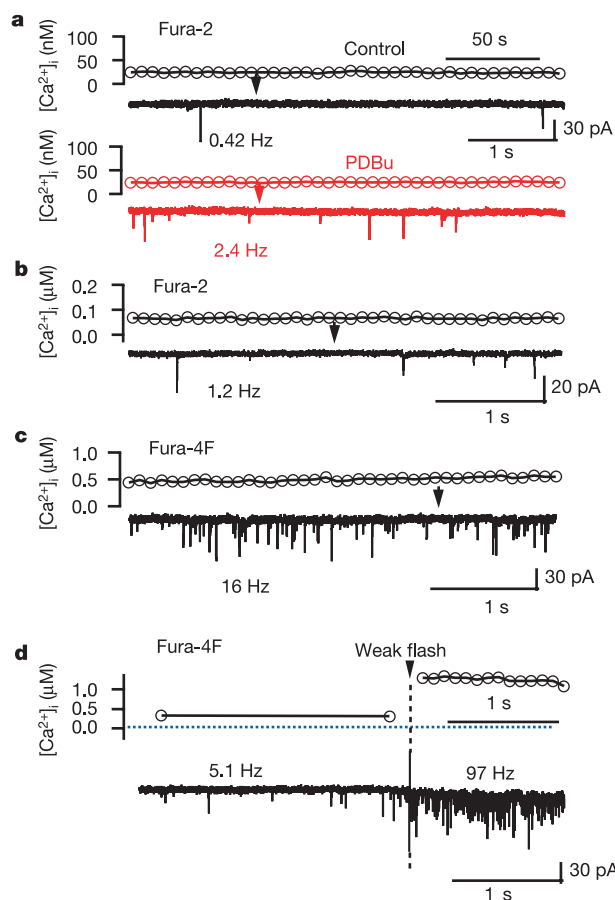


Figure 3 | Transmitter release regulation by presynaptic $[Ca^{2+}]_i$ close to resting values. **a**, Presynaptic $[Ca^{2+}]_i$ (upper trace) and postsynaptic current (lower trace). The presynaptic calyx was pre-loaded with the Ca^{2+} -indicator fura-2. The time scale on the $[Ca^{2+}]_i$ trace applies to **a–c**, and the arrows indicate the times corresponding to the current traces in the lower panels. **b**, Presynaptic $[Ca^{2+}]_i$ (upper trace) and postsynaptic current (lower trace) in a recording in which the calyx was loaded with a solution containing 100 nM free $[Ca^{2+}]_i$. **c**, Same as in **b**, but for a calyx loaded with 500 nM $[Ca^{2+}]_i$. **d**, Presynaptic $[Ca^{2+}]_i$ and postsynaptic current before and after a weak Ca^{2+} -uncaging stimulus.

relative effectiveness of PDBu decreases with increasing $[Ca^{2+}]_i$. With $[Ca^{2+}]_i$ steps to $\sim 10 \mu M$, PDBu does not induce a significant potentiation of peak transmitter release rates (Fig. 2f, rightmost average data point; $P > 0.5$), consistent with the example experiment shown in Fig. 2c. The delay in transmitter release is shortened by PDBu for $[Ca^{2+}]_i$ steps below $8 \mu M$ (Fig. 2h; $P < 0.003$), indicating that phorbol esters slightly increase transmitter release kinetics.

The results in Fig. 2 show that PDBu increases the intracellular Ca^{2+} -sensitivity of vesicle fusion in response to submaximal Ca^{2+} stimuli. To investigate whether this increase in Ca^{2+} -sensitivity is related to the potentiation of spontaneous transmitter release by phorbol esters^{2,4} (Fig. 1c), we related the mEPSC frequency measured in postsynaptic recordings to the $[Ca^{2+}]_i$ in the presynaptic terminal. In a first series of experiments (Fig. 3a), we monitored basal presynaptic $[Ca^{2+}]_i$ after pre-loading calyces with the Ca^{2+} indicator fura-2 during a brief (~ 1 – 2 min) presynaptic whole-cell recording. PDBu increased the frequency of spontaneous mEPSCs from 0.45 ± 0.05 Hz to 1.7 ± 0.3 Hz ($n = 5$ cells), without a significant

change in the basal $[Ca^{2+}]_i$ (from 24 ± 1 nM to 22 ± 1 nM; Fig. 4a, black and red filled squares). The frequency of unstimulated mEPSCs in these recordings was identical to that observed in experiments without any presynaptic recording (0.41 ± 0.10 Hz; $n = 7$ cells; $P = 0.74$, unpaired t -test). In another series of experiments (Fig. 3b, c), presynaptic $[Ca^{2+}]_i$ was clamped to values higher than baseline, by using strongly Ca^{2+} -buffered presynaptic pipette solutions in simultaneous pre- and postsynaptic recordings. A fluorescent Ca^{2+} indicator (fura-2 or fura-4F; $100 \mu M$) measured the effective $[Ca^{2+}]_i$ inside the terminal (see Methods). In the example shown in Fig. 3b, presynaptic $[Ca^{2+}]_i$ was ~ 80 nM, and the rate of transmitter release was 1.2 Hz. In the experiment shown in Fig. 3c, $[Ca^{2+}]_i$ was 500 nM, and the average release activity was much higher (16 Hz). In an additional series of experiments, we applied weak Ca^{2+} -uncaging stimuli that elevated presynaptic $[Ca^{2+}]_i$ to 0.3 – $1.5 \mu M$, and measured the resulting Ca^{2+} -triggered increase in mEPSC frequency (Fig. 3d).

The experiments in Fig. 3 reveal a surprisingly low cooperativity of Ca^{2+} in triggering vesicle release at low $[Ca^{2+}]_i$. We plotted the rates of transmitter release estimated at low $[Ca^{2+}]_i$ (Fig. 3) as a function of $[Ca^{2+}]_i$ (Fig. 4a, squares and open symbols), and also added the data points obtained under control conditions with Ca^{2+} uncaging resulting in higher $[Ca^{2+}]_i$ (Fig. 4a, filled circles). In the double-logarithmic coordinates of this plot, it can be seen that the slope, which corresponds to the exponent in the power relationship between transmitter release and $[Ca^{2+}]_i$, is high at 2 – $8 \mu M$ $[Ca^{2+}]_i$ (slope = 4.2) but becomes progressively smaller at lower $[Ca^{2+}]_i$ (slope = 2.77 at 200 – 700 nM $[Ca^{2+}]_i$ and slope = 0.66 at 20 – 200 nM $[Ca^{2+}]_i$; see blue lines in Fig. 4a).

The observed gradual decrease in Ca^{2+} cooperativity at lower $[Ca^{2+}]_i$ does not agree with predictions from conventional models of Ca^{2+} -activation of vesicle fusion, which assume that vesicle fusion occurs exclusively from a fully occupied Ca^{2+} -sensor^{9,10}. As expected, the prediction of one such model¹⁰ approached its highest slope at low $[Ca^{2+}]_i$, and predicted basal vesicle fusion rates several orders of magnitude lower than the observed ones (Fig. 4a, continuous grey line). We next tested whether the overall Ca^{2+} -dependency of release could be explained by the sum of a steeply Ca^{2+} -dependent release mechanism and Ca^{2+} -independent, spontaneous release, possibly mediated by molecularly different forms of SNARE (soluble NSF-attachment protein receptor) proteins^{17–19}. We therefore added a constant, Ca^{2+} -independent release rate of $0.45 s^{-1}$ to the prediction

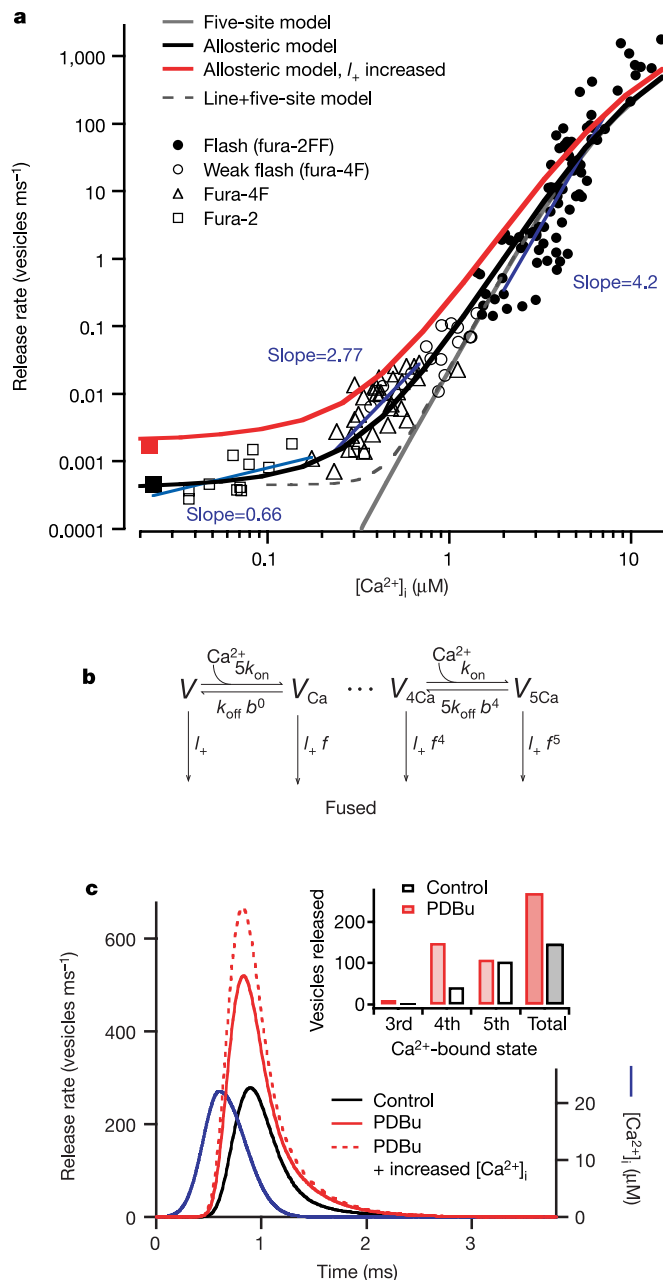


Figure 4 | An allosteric model of Ca^{2+} activation of vesicle fusion. This model explains the Ca^{2+} -sensitivity of transmitter release over a wide range of $[Ca^{2+}]_i$, and explains the effects of PDBu on spontaneous and Ca^{2+} -evoked transmitter release. **a**, Transmitter release rate as a function of $[Ca^{2+}]_i$. The filled squares are average mEPSC frequency under control conditions (black) and after application of PDBu (red) at basal $[Ca^{2+}]_i$ (see Fig. 3a). Data obtained in Ca^{2+} perfusion experiments (open squares, open triangles; Fig. 3b, c) or with weak Ca^{2+} uncaging (open circles; Fig. 3d) are shown. Control condition data points (closed circles) are re-plotted from Fig. 2e. Lines fitted to the logarithmically transformed data and the resulting slopes are shown in blue. The grey line is the prediction of a five-site model for Ca^{2+} -binding and vesicle fusion, with parameters as in ref. 10. The grey dashed line is the prediction of the five-site model to which a Ca^{2+} -independent release rate of $0.45 s^{-1}$ was added. The black and red lines are predictions of a simplified allosteric model (**b**) for two different vesicle fusion rates I_+ (see Methods). **c**, Transmitter release rates predicted by the allosteric model in response to a brief $[Ca^{2+}]_i$ spike (blue trace). Responses for the standard set of parameters (black trace; see Methods) and for a fivefold increase in vesicle fusion rate I_+ (red trace) are shown. The dashed red trace was obtained by further increasing the $[Ca^{2+}]_i$ transient to 110% of the control value, thus accounting for a 10% increase in presynaptic Ca^{2+} current. The inset shows the absolute number of vesicles released from the third, fourth and fifth Ca^{2+} -bound state (see **b**), both under control conditions (black) and with a fivefold increase in vesicle fusion rate (I_+).

of the conventional five-site model, resulting in the dashed grey line (Fig. 4a). This prediction markedly underestimates release rates at 200–800 nM $[Ca^{2+}]_i$, where the observed rates were 9.6 ± 0.9 -fold higher than the prediction ($n = 46$; $P < 0.001$; one-sample t -test). Finally, the data cannot be explained by a highly Ca^{2+} -sensitive sub-pool of vesicles, as observed in chromaffin cells²⁰. This is because the increase in mEPSC frequency in response to a sustained $[Ca^{2+}]_i$ elevation by weak Ca^{2+} uncaging (Fig. 3d) did not show signs of exhaustion, as would be expected for a separate sub-pool of vesicles. Also, adding a single additional sub-pool of vesicles would not explain the rather continuous decrease in Ca^{2+} cooperativity observed below 3 μM $[Ca^{2+}]_i$ (Fig. 4a).

The reduction in Ca^{2+} cooperativity at lower $[Ca^{2+}]_i$ is an expected property of allosteric models, which predict conformational changes of biological molecules in the absence of ligand binding^{21,22}. Allosteric models might also explain how phorbol esters increase the effective Ca^{2+} -sensitivity of vesicle fusion (Fig. 2). We therefore developed an allosteric model of Ca^{2+} -activated vesicle fusion (Fig. 4b), which allows low rates of vesicle fusion in the absence of bound Ca^{2+} (rate constant, $l_+ = 2 \times 10^{-4} s^{-1}$), and in which increasingly higher rates of vesicle fusion are attained when the Ca^{2+} -sensor is more completely occupied (Fig. 4b). This model predicts the rates of transmitter release over the entire range of $[Ca^{2+}]_i$ studied here, from ~ 25 nM to more than 10 μM $[Ca^{2+}]_i$ (Fig. 4a, black line; see Methods for model parameters). The model also accounts for the phorbol-ester-induced potentiation of transmitter release. Assuming that phorbol ester increases the basal fusion 'willingness' (rate constant l_+ in Fig. 4b), an increase in transmitter release at low $[Ca^{2+}]_i$ is predicted (Fig. 4a, compare red and black lines). The increase in l_+ also predicts enhanced fusion rates at intermediate $[Ca^{2+}]_i$ (2–8 μM ; Fig. 4a, red line), with decreased effectiveness of PDBu potentiation at higher $[Ca^{2+}]_i$; again, this is as we observed (Fig. 2e, f). Finally, the model predicts the potentiation of transmitter release evoked by a presynaptic action potential (Fig. 4c). In this simulation, the release response to a brief $[Ca^{2+}]_i$ transient (Fig. 4c, blue trace) was modelled^{9,10}, either for the basal conditions or for a fivefold enhanced rate constant l_+ (Fig. 4c, black and red traces). Taking into account an $\sim 10\%$ increase in presynaptic Ca^{2+} current by PDBu (see Supplementary Fig. 2), a potentiation of transmitter release to 228% of control is predicted (compare the black and the dotted red trace in Fig. 4c), similar as observed for EPSCs evoked by afferent fibre stimulation ($227 \pm 14\%$, Fig. 1b). The model predicts that the potentiation of transmitter release in response to a presynaptic action potential is largely due to increased release out of the fourth Ca^{2+} -bound state, whereas release out of the fully Ca^{2+} -bound state is unaffected (see Fig. 4c, inset).

We have shown that phorbol esters increase transmitter release by enhancing the Ca^{2+} -sensitivity of vesicle fusion, without a significant increase in the size of the readily releasable vesicle pool (Fig. 2d and Supplementary Fig. 2), but with a small contribution from an enhanced presynaptic Ca^{2+} current (see Supplementary Fig. 2). In pituitary cells, phorbol esters increase Ca^{2+} -sensitivity by increasing the vesicle pool depletion time constants²³; however, such an increase is not apparent at the calyx of Held (Fig. 2b). In the framework of a simple allosteric model, phorbol esters enhance the apparent Ca^{2+} -sensitivity by increasing the 'willingness' of vesicle fusion, corresponding to an increase in the rate constant l_+ (Fig. 4b), without a change in the Ca^{2+} -binding properties of the Ca^{2+} -sensor for vesicle fusion. An increased fusion willingness might be caused by lowering the energy barrier for vesicle fusion, a mechanism also implicated in short-term synaptic enhancement and in Ca^{2+} -independent, hypertonicity-evoked transmitter release²⁴. Our allosteric model predicts that an enhanced fusion willingness potentiates the rate of vesicle fusion at baseline $[Ca^{2+}]_i$, as well as the rate of Ca^{2+} -evoked vesicle fusion in an intermediate range of $[Ca^{2+}]_i$ (Fig. 4a, c). Our observations thus explain the link between an increase in the rate of unstimulated transmitter release and a

concomitant increase in transmitter release probability, observed during many forms of presynaptic plasticity^{2,25,26}.

Synaptic vesicle fusion at resting $[Ca^{2+}]_i$ does not appear to be constitutive. Rather, it is regulated by phorbol esters as well as by Ca^{2+} (Fig. 3), although the Ca^{2+} cooperativity is low close to baseline $[Ca^{2+}]_i$ (<1 ; Fig. 4a). Evidence for a low Ca^{2+} cooperativity close to resting $[Ca^{2+}]_i$ has also been obtained at neuromuscular synapses^{27,28}. This low Ca^{2+} cooperativity is functionally relevant, because it prevents a large increase in asynchronous release activity during moderate elevations of residual $[Ca^{2+}]_i$, which would be expected if a steeply Ca^{2+} -dependent mechanism operated close to baseline $[Ca^{2+}]_i$. It has been shown that genetic deletion of the SNARE proteins synaptobrevin^{17,18} or SNAP-25 (ref. 19) affect Ca^{2+} -evoked release more strongly than unstimulated release or release stimulated by hypertonicity. This is consistent with our observations, if our simplified allosteric model is assumed to describe the kinetic responses of a complex of several proteins (including SNAREs and synaptotagmins) in synaptic vesicle fusion^{7,8}. Removing one essential protein might affect the increase in vesicle fusion rate upon Ca^{2+} -binding (rate constant f in Fig. 4b) to a greater degree than the basal fusion 'willingness' (rate constant l_+ in Fig. 4b). In our view, spontaneous and Ca^{2+} -evoked transmitter release thus emerge from a fine-tuned molecular complex that mediates vesicle fusion at synapses.

METHODS

Slice preparation and electrophysiology. Transverse brainstem slices (200 μm) containing the medial nucleus of the trapezoid body (MNTB) were prepared from 8–11-day-old rat brains using a vibratome. Simultaneous whole-cell patch-clamp recordings from a postsynaptic MNTB principal cell and its presynaptic calyx of Held were made under visual control, using an EPC10/2 double amplifier (HEKA). Both cells were voltage-clamped at -80 mV, and series resistance was compensated up to 90%. The remaining error in the postsynaptic currents was corrected off-line. The extracellular solution was a bicarbonate buffered Ringer solution (for example, see ref. 11) with 2 mM $CaCl_2$ and 1 mM $MgCl_2$. For paired recordings, 0.5 μM tetrodotoxin, 10 mM tetraethylammonium chloride (TEA), 50 μM D-2-amino-5-phosphonovaleric acid and 100 μM cyclothiazide (CTZ) were added to the solution. For experiments with strong Ca^{2+} -uncaging stimuli (Fig. 2c, d), 2 mM γ -D-glutamyl-glycine (γ -DGG) was also present to prevent AMPA-receptor saturation. The pipette (intracellular) solution contained caesium-gluconate¹¹, with 0.2 and 5 mM EGTA for pre- and postsynaptic recordings, respectively. Experiments were done at room temperature (21 – $24^\circ C$). In some experiments, EPSCs were evoked by afferent fibre stimulation without simultaneous presynaptic recording (Fig. 1).

Ca^{2+} uncaging and Ca^{2+} imaging. For Ca^{2+} uncaging, the presynaptic pipette (intracellular) solution contained (mM): 130 caesium gluconate, 20 TEA-Cl, 20 HEPES, 5 Na_2ATP , 0.3 Na_2GTP , 0.1 fura-2FF (Teflabs), 1 DM-nitrophen (Calbiochem), 0.85 $CaCl_2$ and 0.5 $MgCl_2$. Ca^{2+} uncaging was done with a flash lamp (Rapp Optoelektronik), and $[Ca^{2+}]_i$ was measured by imaging the fluorescence of fura-2FF with a slow-scan CCD camera (TILL-photronics) at excitation wavelengths of 350 and 380 nm. $[Ca^{2+}]_i$ was calculated from background-corrected fluorescence ratios, using calibration procedures as described¹¹.

For clamping the presynaptic $[Ca^{2+}]_i$ in the range of 0.05–1 μM (Fig. 3b, c), the presynaptic pipette solution contained 20 mM EGTA, to which different concentrations of $CaCl_2$ were added to yield nominal free $[Ca^{2+}]_i$ of 50, 100, 300 and 600 nM. The effective presynaptic $[Ca^{2+}]_i$ was measured with fura-2 or fura-4F added to the intracellular solution (100 μM).

Data analysis and simulations. Analysis routines were written in IgorPro (Wavemetrics). Miniature EPSCs were detected using a template matching routine. Transmitter release rates were calculated by deconvolving the measured EPSCs with an idealized mEPSC waveform with double-exponential decay^{10,16}. The deconvolution assumes that EPSCs arise from linearly summed mEPSCs with amplitudes of 30 pA (for CTZ) and 15 pA (for CTZ and γ -DGG), and an additional current component generated by glutamate spillover, which was subtracted. The minimal synaptic delays (Fig. 2g) are expressed as the time between the trigger of the flash lamp and the occurrence of five quanta in integrated release-rate traces. Cumulative release rates were fitted with double-exponential functions. An exponential with the fitted parameters for the fast component (dashed lines in Fig. 2b, d) was subtracted from the data trace to yield the isolated slow component (dotted lines in Fig. 2b, d). Unless stated

otherwise, data are reported as mean $\pm$ s.e.m., and statistical significance was determined by paired *t*-tests.

The relative change in transmitter release rates induced by PDBu (Fig. 2f) was calculated for individual cells as follows¹¹. First, the line fit to the logarithmically-transformed control data set (Fig. 2e, black line) was shifted on the *x* axis ($\log([Ca^{2+}]_i)$) to fit the control data points for a given recording. Then, the difference in *y* value ($\log(\text{release rate})$) between each data point obtained in PDBu and the shifted fit line was read out. In this way, possible differences in $[Ca^{2+}]_i$ are taken into account. The relative change in delays (Fig. 2h) was analysed in the same way.

Ca^{2+} -dependent vesicle fusion rates were simulated using a conventional five-site model¹⁰ and a simplified allosteric model (Fig. 4b). Both models were solved numerically. Release was assumed to occur from a homogeneous pool of 2,000 vesicles. The simulations according to the five-site model (Fig. 4a, continuous grey line) were made with the parameters given in ref. 10. Parameter values for the simplified allosteric model (Fig. 4b) were constrained as follows. First, the factor determining the increase in vesicle fusion rate upon Ca^{2+} -binding, *f*, was set to 31.3, giving $I_+ \times f^5 = 6,000 \text{ s}^{-1}$, identical to the fusion rate (γ) of the five-site model in ref. 10. Then, k_{on} ($1 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$), k_{off} ($4,000 \text{ s}^{-1}$) and the cooperativity factor, *b* (0.5) were set such that the predictions of the two models at $[Ca^{2+}]_i > 7 \mu\text{M}$ were similar (Fig. 4a). Finally, the rate constant $I_+(2 \times 10^{-4} \text{ s}^{-1})$ was set to fit the vesicle fusion rate at low $[Ca^{2+}]_i$. For predicting release in the presence of phorbol ester (red traces in Fig. 4a, c), I_+ was increased fivefold, with all other parameters unchanged.

Received 13 January; accepted 22 March 2005.

- Malenka, R. C., Madison, D. V. & Nicoll, R. A. Potentiation of synaptic transmission in the hippocampus by phorbol ester. *Nature* **321**, 175–177 (1986).
- Shapira, R., Silberberg, S. D., Ginsburg, S. & Rahamimoff, R. Activation of protein kinase C augments evoked transmitter release. *Nature* **325**, 58–60 (1987).
- Yawo, H. Protein kinase C potentiates transmitter release from the chick ciliary presynaptic terminal by increasing the exocytotic fusion probability. *J. Physiol. (Lond.)* **515**, 169–180 (1999).
- Hori, T., Takai, Y. & Takahashi, T. Presynaptic mechanism for phorbol ester-induced synaptic potentiation. *J. Neurosci.* **19**, 7262–7267 (1999).
- Rhee, J.-S. *et al.* β phorbol ester- and diacylglycerol-induced augmentation of transmitter release is mediated by munc13s and not by PKCs. *Cell* **108**, 121–133 (2002).
- Silinsky, E. M. & Searl, T. J. Phorbol esters and neurotransmitter release: More than just protein kinase C? *Br. J. Pharmacol.* **138**, 1191–1201 (2003).
- Chen, Y. A. & Scheller, R. H. SNARE-mediated membrane fusion. *Nature Rev. Mol. Cell Biol.* **2**, 98–106 (2001).
- Jahn, R., Lang, T. & Südhof, T. C. Membrane fusion. *Cell* **112**, 519–533 (2003).
- Bollmann, J., Sakmann, B. & Borst, J. Calcium sensitivity of glutamate release in a calyx-type terminal. *Science* **289**, 953–957 (2000).
- Schneggenburger, R. & Neher, E. Intracellular calcium dependence of transmitter release rates at a fast central synapse. *Nature* **406**, 889–893 (2000).
- Felmy, F., Neher, E. & Schneggenburger, R. Probing the intracellular calcium sensitivity of transmitter release during synaptic facilitation. *Neuron* **37**, 801–811 (2003).
- Wu, X. S. & Wu, L. G. Protein kinase C increases the apparent affinity of the release machinery to Ca^{2+} by enhancing the release machinery downstream of the Ca^{2+} sensor. *J. Neurosci.* **21**, 7928–7936 (2001).
- Ishikawa, T., Sahara, Y. & Takahashi, T. A single packet of transmitter does not saturate postsynaptic glutamate receptors. *Neuron* **34**, 613–621 (2002).
- Gillis, K. D., Mößner, R. & Neher, E. Protein kinase C enhances exocytosis from chromaffin cells by increasing the size of the readily releasable pool of secretory granules. *Neuron* **16**, 1209–1220 (1996).
- Stevens, C. F. & Sullivan, J. M. Regulation of the readily releasable vesicle pool by protein kinase C. *Neuron* **21**, 885–893 (1998).
- Sakaba, T. & Neher, E. Calmodulin mediates rapid recruitment of fast-releasing synaptic vesicles at a calyx-type synapse. *Neuron* **32**, 1119–1131 (2001).
- Deitcher, D. L. *et al.* Distinct requirements for evoked and spontaneous release of neurotransmitter are revealed by mutations in the *Drosophila* gene *neuronal-synaptobrevin*. *J. Neurosci.* **18**, 2028–2039 (1998).
- Schoch, S. *et al.* SNARE function analyzed in synaptobrevin/VAMP knockout mice. *Science* **294**, 1117–1122 (2001).
- Washbourne, P. *et al.* Genetic ablation of the t-SNARE SNAP-25 distinguishes mechanisms of neuroexocytosis. *Nature Neurosci.* **5**, 19–26 (2002).
- Yang, Y., Udayasankar, S., Dunning, J., Chen, P. & Gillis, K. D. A highly Ca^{2+} -sensitive pool of vesicles is regulated by protein kinase C in adrenal chromaffin cells. *Proc. Natl Acad. Sci. USA* **99**, 17060–17065 (2002).
- Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
- Li, J., Zagotta, W. N. & Lester, H. A. Cyclic nucleotide-gated channels: structural basis of ligand efficacy and allosteric modulation. *Q. Rev. Biophys.* **30**, 177–193 (1997).
- Zhu, H., Hille, B. & Xu, T. Sensitization of regulated exocytosis by protein kinase C. *Proc. Natl Acad. Sci. USA* **99**, 17055–17059 (2002).
- Stevens, C. F. & Wesseling, J. F. Augmentation is a potentiation of the exocytotic process. *Neuron* **22**, 139–146 (1999).
- Scanziani, M., Capogna, M., Gähwiler, B. H. & Thompson, S. M. Presynaptic inhibition of miniature excitatory synaptic currents by baclofen and adenosine in the hippocampus. *Neuron* **9**, 919–927 (1992).
- Piedras-Rentería, E. S. *et al.* Presynaptic homeostasis at CNS nerve terminals compensates for lack of a key Ca^{2+} entry pathway. *Proc. Natl Acad. Sci. USA* **101**, 3609–3614 (2004).
- Delaney, K. R. & Tank, D. W. A quantitative measurement of the dependence of short-term synaptic enhancement on presynaptic residual calcium. *J. Neurosci.* **14**, 5885–5902 (1994).
- Angleton, J. K. & Betz, W. J. Intraterminal Ca^{2+} and spontaneous transmitter release at the frog neuromuscular junction. *J. Neurophysiol.* **85**, 287–294 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Neher for discussions, and P. Ascher, F. Felmy, E. Neher, T. Sakaba and M. Wölfel for comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (DFG). R.S. is a Heisenberg fellow of the DFG.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.S. (rschneg@gwdg.de).

LETTERS

Vascular respiratory uncoupling increases blood pressure and atherosclerosis

Carlos Bernal-Mizrachi^{1*}, Allison C. Gates^{1*}, Sherry Weng¹, Takuji Imamura², Russell H. Knutsen³, Pascual DeSantis¹, Trey Coleman¹, R. Reid Townsend¹, Louis J. Muglia² & Clay F. Semenkovich^{1,3}

The observations that atherosclerosis often occurs in non-smokers without elevated levels of low-density lipoprotein cholesterol, and that most atherosclerosis loci so far identified in mice do not affect systemic risk factors associated with atherosclerosis¹, suggest that as-yet-unidentified mechanisms must contribute to vascular disease. Arterial walls undergo regional disturbances of metabolism² that include the uncoupling of respiration and oxidative phosphorylation, a process that occurs to some extent in all cells and may be characteristic of blood vessels being predisposed to the development of atherosclerosis³. To test the hypothesis that inefficient metabolism in blood vessels promotes vascular disease, we generated mice with doxycycline-inducible expression of uncoupling protein-1 (UCP1) in the artery wall. Here we show that UCP1 expression in aortic smooth muscle cells causes hypertension and increases dietary atherosclerosis without affecting cholesterol levels. UCP1 expression also increases superoxide production and decreases the availability of nitric oxide, evidence of oxidative stress. These results provide proof of principle that inefficient metabolism in blood vessels can cause vascular disease.

Oxidative stress is implicated in atherosclerosis, but energy metabolism in the vascular wall, a primary site of reactive oxygen species (ROS) production, is poorly understood. Mitochondrial electron transport in higher organisms accounts for most ROS production⁴. Mitochondrial respiration produces energy in the form of ATP, resulting from ADP phosphorylation, as well as ROS, as electrons at complex I and III react with molecular oxygen to form superoxide⁵. Uncoupling proteins (inner mitochondrial membrane anion transporters) allow protons to leak back into the mitochondrial matrix, thereby decreasing the potential energy available for ADP phosphorylation and ROS generation. They also increase respiration, which is not thought to accelerate superoxide production in the setting of low protonmotive force⁶. Superoxide activates uncoupling proteins⁷, which have been postulated to limit further superoxide generation by dissipating protonmotive force and thereby thwarting oxidative stress. Consistent with this idea, uncoupling decreases glucose-induced ROS formation and interrupts pathways associated with vascular damage in cultured endothelial cells⁸. UCP2 in macrophages was reported to decrease ROS and atherosclerosis⁹. These results seem to be in conflict with evidence that inefficient vascular metabolism can be detrimental. Uncoupling agents produce smooth muscle contraction¹⁰, a hallmark of hypertension. Respiratory uncoupling is increased in the aortae of pigeons susceptible to atherosclerosis³, suggesting that the efficiency of vascular wall energy metabolism could be a determinant of atherosclerotic lesion development.

Smooth muscle cells are a principal source of ROS in the

vasculature⁴. To determine whether the degree of respiratory coupling in the vasculature is involved in vascular disease, we generated mice with doxycycline-inducible expression of UCP1 restricted to aortic smooth muscle cells. Transgenic mice carrying the reverse tetracycline transactivator (*rtTA*) driven by the -441 SM22 α promoter (*SM22-rtTA* mice) were mated with mice transgenic for a tetracycline responsive element (TRE)-UCP1 minigene (*TRE-Ucp1* mice), yielding *SM22-TRE-Ucp1* mice. Ten days after treating these mice with doxycycline (2 mg ml⁻¹) in sucrose-containing drinking water, aortic *Ucp1* messenger RNA expression was induced by approximately 12-fold compared with mice drinking sucrose-containing water alone (Fig. 1a). There was no doxycycline-dependent induction of *Ucp1* gene expression in the aortae of *SM22-rtTA* mice (lacking a TRE target) or *TRE-Ucp1* mice (lacking an antibiotic-inducible transactivator; data not shown).

UCP1 protein was also induced in aortae of doxycycline-treated mice (Fig. 1b, top panel, lanes 1–3) as compared with aortae of mice not receiving doxycycline (Fig. 1b, bottom panel, lanes 1–3). As expected, UCP1 was detected in brown fat regardless of doxycycline treatment (lane 10) but not in other tissues (lanes 4–9). Repeat western blotting of smooth muscle-containing tissues (uterus, urethra, stomach, colon, bladder) showed no evidence of doxycycline-dependent UCP1 expression (data not shown), confirming that

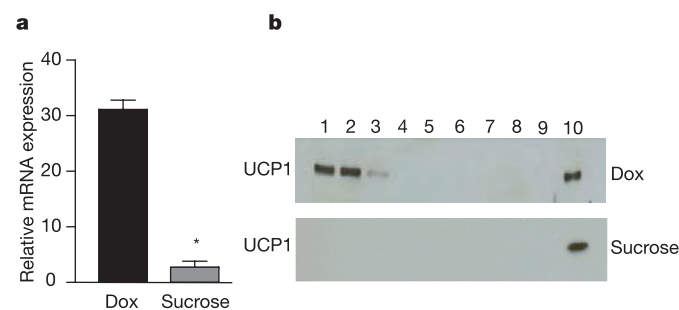


Figure 1 | Inducible expression of vascular wall UCP1. **a**, *SM22-TRE-Ucp1* mice were treated with doxycycline (Dox; 2 mg ml⁻¹) in sucrose or sucrose alone for 10 days, followed by determination of *Ucp1* mRNA levels (normalized to *Gapd*) in the aorta. Results are mean $\pm$ s.e.m. for four mice per condition. The asterisk indicates $P < 0.01$. **b**, Tissue-specific expression of UCP1 protein after 10 days of doxycycline treatment (top panel) or control (bottom panel). Lanes were probed with an anti-UCP1 antibody after loading as follows: 1, aortic arch; 2, thoracic aorta; 3, abdominal aorta; 4, quadriceps muscle; 5, heart; 6, bladder; 7, colon; 8, kidney; 9, lung; 10, brown fat.

¹Department of Medicine, Division of Endocrinology, Metabolism & Lipid Research, ²Department of Pediatrics, and ³Department of Cell Biology & Physiology, Washington University School of Medicine, St Louis, Missouri 63110, USA.

*These authors contributed equally to this work.

the -441 *SM22* promoter fragment drives transcription limited to arterial smooth muscle¹¹. Aortic induction of UCP1 expression did not affect body weight or serum chemistries in *SM22-TRE-Ucp1* mice (Supplementary Fig. 1).

Doxycycline induction of aortic UCP1 expression increased tail-cuff systolic blood pressure by 46 mm Hg ($P = 0.002$) and diastolic pressure by 21 mm Hg (Fig. 2a, $P = 0.005$, black bars). Blood pressure elevations with doxycycline treatment were confirmed with implanted telemetry sensors (Supplementary Fig. 2a) and invasively under anaesthesia (data not shown). Blood pressure returned to baseline in the same animals 10 days after removing doxycycline from the drinking water (grey bars). Plasma renin activity—a critical determinant of renal sodium handling and blood pressure control—nearly tripled in doxycycline-treated mice and returned to baseline values after discontinuing doxycycline treatment (Fig. 2b). Urinary sodium excretion decreased in the presence of doxycycline (Fig. 2c), consistent with activation of the renin–angiotensin–aldosterone system. Urinary excretion of norepinephrine, a marker of sympathetic activation, was not altered by UCP1 induction (data not shown).

To study the effects of UCP1 induction on vascular disease, mice carrying the *SM22-rtTA* and *TRE-Ucp1* transgenes were bred with apolipoprotein E (*Apoe*)-null mice, a robust atherosclerosis model. Doxycycline induction of aortic UCP1 was documented in *SM22-TRE-Ucp1 Apoe*^{-/-} mice (data not shown), and these animals were fed a Western diet (see Methods) (Fig. 2d). Tail-cuff blood pressures were higher ($P = 0.008$ for systolic, $P = 0.04$ for diastolic)

in the presence (Fig. 2d, black bars) as compared with the absence of doxycycline (grey bars). In *SM22-TRE-Ucp1 Apoe*^{-/-} mice fed a chow diet (low fat, no added cholesterol), mean arterial pressure determined by telemetry was increased by doxycycline treatment (Supplementary Fig. 2b). Doxycycline increased plasma renin activity (Fig. 2e) and decreased urinary sodium excretion (Fig. 2f) in the *Apoe* null model. The effects of doxycycline on blood pressure required induction of *Ucp1* gene expression, because doxycycline treatment of either *SM22-rtTA Apoe*^{-/-} mice (animals expressing the transactivator without a target) or *TRE-Ucp1 Apoe*^{-/-} mice (carrying the target without a transactivator) had no effect on blood pressure (Fig. 2g, i) or renin (Fig. 2h, j).

After 6 weeks on the Western diet, doxycycline induction of UCP1 in the aortae of *SM22-TRE-Ucp1 Apoe*^{-/-} mice (Fig. 3a, filled symbols) increased atherosclerosis as compared with littermate mice not treated with doxycycline (Fig. 3a, open symbols). UCP1 induction in chow-fed *SM22-TRE-Ucp1 Apoe*^{-/-} mice did not increase lesions after 6 weeks (Supplementary Fig. 3) but did increase blood pressure (Supplementary Fig. 2b).

Doxycycline alone may have several effects on vascular biology¹², but total (Fig. 3d) as well as lipoprotein (Fig. 3d inset) cholesterol levels were unaffected by doxycycline treatment in these animals, and doxycycline treatment did not increase atherosclerosis in either of the controls: *TRE-Ucp1 Apoe*^{-/-} mice (Fig. 3b) and *SM22-rtTA Apoe*^{-/-} mice (Fig. 3c). In fact, doxycycline modestly decreased diet-induced atherosclerosis in both controls, probably because of lower serum cholesterol levels (Fig. 3e, f). Induction of UCP1 in the aortae of

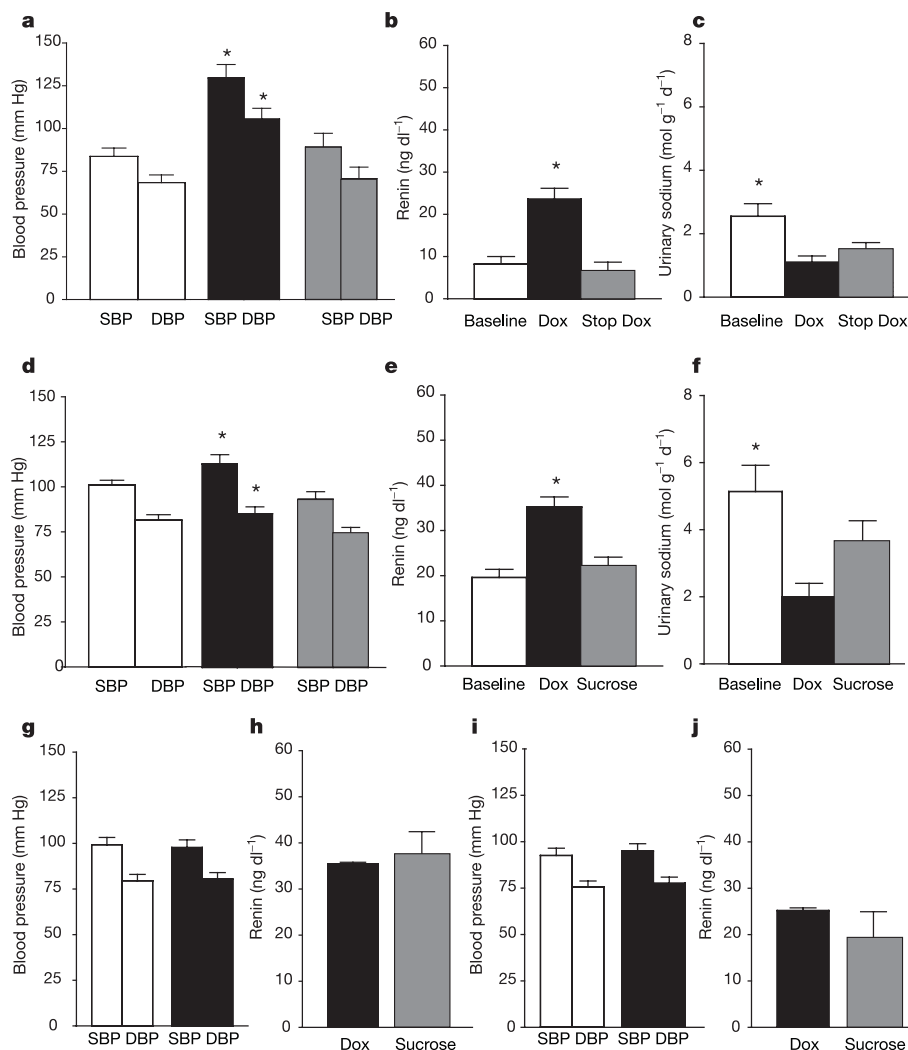


Figure 2 | Effects of vascular UCP1 expression on blood pressure, plasma renin activity and urinary sodium. a–c, Data from *SM22-TRE-Ucp1* mice with wild-type apolipoprotein E alleles; **d–f**, data from *SM22-TRE-Ucp1 Apoe*^{-/-} mice; **g, h**, data from *TRE-Ucp1 Apoe*^{-/-} mice; **i, j**, data from *SM22-rtTA Apoe*^{-/-} mice. **a**, Systolic blood pressure (SBP) and diastolic blood pressure (DBP) at baseline (white bars, $n = 10$), after 7 days of doxycycline treatment (black bars, $n = 10$) and 10 days after doxycycline treatment was discontinued (grey bars, $n = 5$). Asterisk indicates $P < 0.01$ versus baseline and discontinuation. **b**, Plasma renin activity at baseline (white bar), after 7 days of doxycycline treatment (black bar) and 10 days after discontinuation (grey bar). Asterisk indicates $P < 0.01$ versus baseline and discontinuation. **c**, Urinary sodium excretion for mice of **a** and **b**. Asterisk indicates $P = 0.011$ versus doxycycline. **d**, Blood pressures for *SM22-TRE-Ucp1 Apoe*^{-/-} mice at baseline (white bars, $n = 20$) and after 6 weeks on Western diet with (black bars, $n = 10$) or without doxycycline (grey bars, $n = 10$). Asterisk indicates $P < 0.01$ versus baseline and no doxycycline treatment for SBP. Asterisk indicates $P < 0.05$ versus no doxycycline treatment for DBP. **e**, Plasma renin activity for mice of **d**. Asterisk indicates $P < 0.01$ versus other conditions. **f**, Urinary sodium excretion for animals of **d**. Asterisk indicates $P < 0.01$ versus doxycycline. **g**, Blood pressures for *TRE-Ucp1 Apoe*^{-/-} mice at baseline (white bars) and after 6 weeks on Western diet with doxycycline (black bars). **h**, Plasma renin activity for mice of **g**. **i**, Blood pressures for *SM22-rtTA Apoe*^{-/-} mice at baseline (white bars) and after 6 weeks on Western diet with doxycycline (black bars). **j**, Plasma renin activity for mice of **i**. All results are mean $\pm$ s.e.m.

SM22-TRE-Ucp1 Apoe^{-/-} mice did not affect weight gain during Western diet feeding, adiposity, whole-body oxygen consumption, glucose tolerance (Supplementary Fig. 4a–d), serum triglycerides or free fatty acids (data not shown). The paucity of systemic metabolic alterations in this model is consistent with restricted expression of UCP1 to a small volume of tissue. Plasma renin activity and blood pressure were increased by doxycycline treatment in *SM22-TRE-Ucp1 Apoe^{-/-}* mice (Fig. 2d–f). We were unable to detect any increase in renal artery UCP1 expression with doxycycline treatment, although UCP1 was induced after doxycycline treatment in femoral and carotid arteries (Supplementary Fig. 5). Periaortic brown fat is present in humans¹³ and rodents¹⁴ in sufficient amounts to alter the temperature of aortic blood after sympathetic stimulation. Low levels of *Ucp1* mRNA that did not change with doxycycline treatment were detected in the aortae of *TRE-Ucp1 Apoe^{-/-}* and *SM22-rtTA Apoe^{-/-}* mice (data not shown).

To implicate the effects of oxidative stress in the vascular phenotype of these animals, superoxide production was measured using aortae from young mice (age 8 weeks) with no atherosclerosis treated with doxycycline for 8 days. Superoxide production was increased at the aortic arch and thoracic aorta of doxycycline-treated mice (Fig. 4a). Consistent with increased oxidative stress in the setting of UCP1 expression, several proteins known to be decreased by ROS were identified by differential proteomic analysis of doxycycline-treated aortae (Fig. 4b). Aconitase 2, α -enolase and actin undergo age-dependent oxidative modification¹⁵. ROS increase smooth muscle cell differentiation and decrease expression of α -actin protein¹⁶. Protein disulphide isomerase and DNA K chaperone can be oxidatively modified during the process of interacting with oxidatively damaged proteins^{17,18}. Aconitase is readily inactivated by superoxide¹⁹. Aortic aconitase enzyme activity was decreased in doxycycline-treated mice (Fig. 4c), confirming increased superoxide

production (Fig. 4a) and decreased aconitase protein (Fig. 4b). Increased superoxide production can decrease the biological availability of nitric oxide (NO), causing increased blood pressure. Intravenous infusion of the nonspecific nitric oxide synthase inhibitor L-NAME increases blood pressure. L-NAME infusion in doxycycline-treated *SM22-TRE-Ucp1* mice resulted in a 33% smaller increase in systolic blood pressure as compared with control mice (Fig. 4d), consistent with less biologically available NO in the setting of doxycycline treatment. Superoxide can react with NO to yield peroxynitrite, which can modify the amino acid tyrosine. More nitrotyrosine staining was detected in aortae of mice with no atherosclerosis treated with doxycycline for 8 days as compared with mice that did not receive doxycycline (Fig. 4e), providing evidence of peroxynitrite production resulting from the reaction between superoxide and NO in aortae expressing UCP1. As additional evidence for the role of oxidative stress in the phenotype, administration of the antioxidant tempol for 5 days reversed the UCP1-induced blood pressure elevation in doxycycline-treated *SM22-TRE-Ucp1 Apoe^{-/-}* mice (Supplementary Fig. 6).

Atherosclerotic lesions in doxycycline-treated *SM22-TRE-Ucp1 Apoe^{-/-}* mice were histologically similar to those in animals not treated with doxycycline. Most cells were macrophages, as determined by staining with a macrophage-specific marker (data not with no atherosclerosis were modestly but not significantly decreased as compared with control aortae (doxycycline treatment: $0.25 \pm 0.05 \mu\text{g ATP per mg protein}$; no doxycycline: $0.31 \pm 0.09 \mu\text{g ATP per mg protein}$; $P = 0.51$). ADP/ATP ratios were also unaffected by doxycycline. Aortic oxygen consumption in mice treated the same way was modestly but not significantly increased (doxycycline treatment: $6.1 \pm 1.2 \text{ nmol min}^{-1} \text{ mg}^{-1}$; no doxycycline: $4.4 \pm 1.8 \text{ nmol min}^{-1} \text{ mg}^{-1}$; $P = 0.43$).

Atherosclerosis is generally thought to be a chronic inflammatory

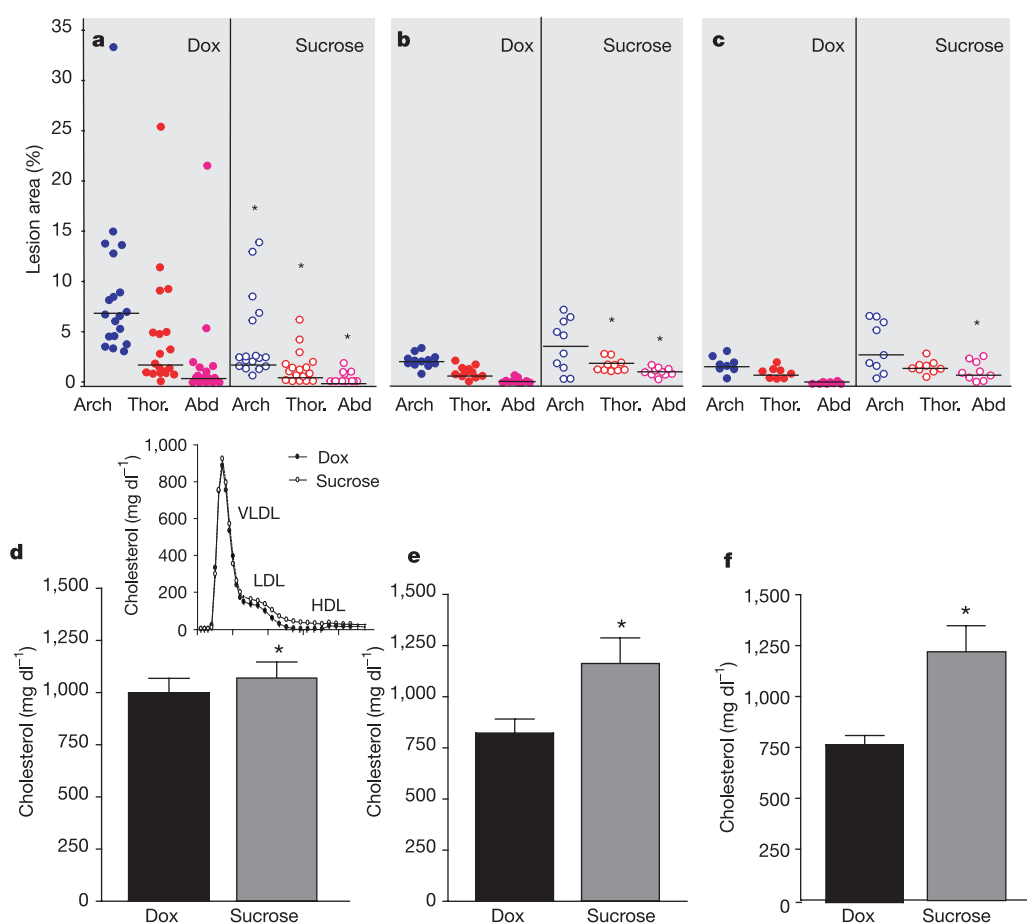


Figure 3 | Effects of vascular UCP1 expression on diet-induced atherosclerosis and serum cholesterol. **a–f**, *SM22-TRE-Ucp1 Apoe^{-/-}* mice (**a**, **d**) in the presence or absence of doxycycline were compared to control mice *TRE-Ucp1 Apoe^{-/-}* (**b**, **e**) and *SM22-rtTA Apoe^{-/-}* (**c**, **f**). Asterisks indicate $P = 0.02$ (arch), $P = 0.008$ (thoracic aorta; Thor.) and $P = 0.02$ (abdominal; Abd) in **a** (Mann–Whitney *U*-test); $P = 0.03$ (thoracic aorta) and $P = 0.0002$ (abdominal) in **b**; $P = 0.0003$ in **c**; and $P = 0.0001$ and $P = 0.037$ in **e**, **f** (unpaired two-tailed *t*-test). Error bars indicate s.e.m. Inset in **d** shows fasting lipoproteins in *SM22-TRE-Ucp1 Apoe^{-/-}* mice. The *x* axis for the inset indicates fraction number. HDL, high-density lipoprotein; LDL, low-density lipoprotein; VLDL, very low-density lipoprotein.

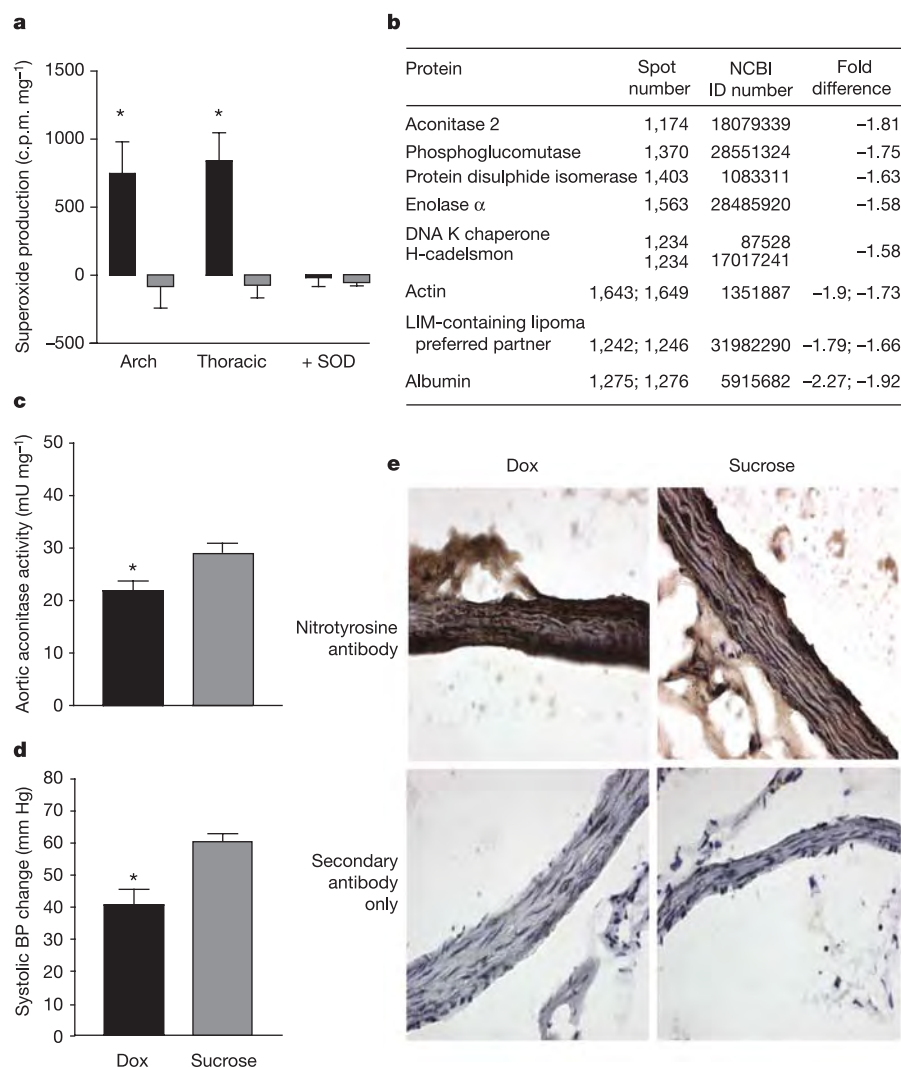


Figure 4 | Involvement of oxidative stress in the vascular phenotype after UCPI

induction. a, Superoxide production. *SM22-TRE-Ucp1 Apoe*^{-/-} mice (three per condition) received doxycycline (black bars) or sucrose alone (grey bars) for 8 days to induce UCPI expression in the vasculature. Superoxide was assayed in aortic slices using lucigenin in the presence (+SOD) or absence of PEG-superoxide dismutase. c.p.m., counts per minute. **b**, Proteomic analysis. UCPI-induced changes in aortic proteins were identified by fluorescent two-dimensional difference gel electrophoresis and mass spectrometry. Proteins were extracted from the aortae of *SM22-TRE-Ucp1 Apoe*^{-/-} mice fed chow and treated with ($n = 3$) or without ($n = 3$) doxycycline for 8 days. Some proteins were detected in more than one spot. **c**, Aconitase enzyme activity. Aortae were assayed from *SM22-TRE-Ucp1 Apoe*^{-/-} mice (four per condition) treated with or without doxycycline for 8 days. **d**, Response to L-NAME infusion. Blood pressure was determined invasively in doxycycline-treated (black bar, $n = 4$) and sucrose-treated (grey bar, $n = 4$) *SM22-TRE-Ucp1* mice after infusion of L-NAME. Values represent mean increase in SBP with L-NAME infusion. Asterisk indicates $P < 0.05$. Results were confirmed in mice of the *Apoe*^{-/-} background (data not shown). **e**, Nitrotyrosine immunohistology. Representative nitrotyrosine staining in *SM22-TRE-Ucp1 Apoe*^{-/-} mice with no atherosclerosis after doxycycline or sucrose treatment for 8 days. Similar results were seen in three animals per condition. For **a**, **c** and **d**, results are mean $\pm$ s.e.m.

disease resulting from vascular injury. However, one of the earliest signs of atherosclerosis is the development of abnormal mitochondria in smooth muscle cells²⁰, suggesting that mitochondrial dysfunction in these cells, the predominant cell type in atherosclerosis-susceptible vessels, triggers disease. Our data provide direct evidence that promoting inefficient mitochondrial metabolism in smooth muscle cells results in hypertension and accelerated atherosclerosis. Increased blood pressure over the 6-week time course of these experiments does not explain the increase in atherosclerosis, because chow-fed mice treated with doxycycline had increased blood pressure (Supplementary Fig. 2b) without an effect on lesions (Supplementary Fig. 3). Hypertension does not consistently modulate atherosclerosis in mice^{21,22}.

Previous data support the idea that the general process of respiratory uncoupling may be involved in atherogenesis. Arteries have marginal oxygenation²³. Hypoxia decreases the respiratory control ratio²⁴; that is, the ratio of phosphorylating (state 3) to non-phosphorylating (state 4) respiration. Uncoupled respiration precedes atherosclerosis at lesion-prone aortic sites in white carneau pigeons but not at the same sites in show racer pigeons (an atherosclerosis-resistant strain), and not at aortic sites that do not develop lesions in either strain³. Disease-free aortae have abundant concentrations of the essential fatty acid linoleate, but fatty streaks are deficient in essential fatty acids^{25,26}. Essential fatty acid deficiency promotes respiratory uncoupling^{27,28} and atherosclerosis²⁹.

Oxidative stress can be defined as increased generation of ROS and

decreased biological availability of NO, both seen with smooth muscle expression of UCPI. These results were unexpected. High protonmotive force, which occurs when both ATP consumption and ADP availability are low, is thought to determine ROS production. However, vigorous exercise also enhances ROS production in skeletal muscle³⁰, establishing precedence for increased ROS production despite extensive ATP consumption and ADP availability.

Thus, mild respiratory uncoupling—a form of mitochondrial dysfunction—in blood vessels of mice increases oxidative stress, blood pressure and dietary atherosclerosis. These data support the notion that local disturbances of metabolism in the arterial wall cause vascular disease. Pharmacological or nutritional strategies to enhance efficient metabolism in the vasculature could represent new approaches for treating atherosclerosis.

METHODS

Animals. For the *SM22-rtTA* transgenic line, the cytomegalovirus (CMV) promoter region in the vector pIRES2-EGFP (Clontech) was replaced with the murine smooth muscle SM22 α promoter region from -441 through to +40 base pairs relative to the start site of transcription. The *rtTA* coding region was then inserted into the polylinker downstream of the SM22 α promoter as a blunt end *EcoRI*-*BamHI* fragment. To generate the *TRE-Ucp1* transgenic line, we used pUCP1, which responds to the doxycycline-rtTA complex in cultured cells (see Supplementary Information and references therein). For both constructs, sequences were separated from vector DNA and C57BL/6 $\times$ CBA hybrid embryos were injected. Single and double transgenic animals were born in the expected mendelian frequencies. Mice with a single transgene were bred with

Apoe-null mice in the C57BL/6 background, and then these animals were bred with each other to generate double transgenic mice with apolipoprotein E deficiency. Although these animals were of mixed genetic background, all experiments were performed with littermates. To address the possibility that observations were due to the integration of transgenes near an atherosclerosis susceptibility locus, control atherosclerosis experiments were conducted using each individual transgene in the *Apoe*-null model (Fig. 3b, c). To induce UCP1 expression, mice were provided with fresh drinking water containing 2 mg ml⁻¹ doxycycline and 5% (w/v) sucrose, whereas controls received a 5% sucrose solution without doxycycline. The Washington University Animal Studies Committee approved these studies.

Blood pressure. Non-invasive determinations were performed in conscious mice by tail cuff (see Supplementary Information and references therein). Telemetry determinations were performed using a system from Data Science International. Vascular responses to inhibition of nitric oxide synthase were recorded invasively (PowerLab/8SP, AD Instruments) after the intravenous infusion of 5 mg kg⁻¹ L-NAME.

Atherosclerosis. At the age of 8 weeks, animals in the *Apoe*-null background were started on a Western-type diet with 0.15% cholesterol, which provides 42% calories as fat (TD 88137); the diet was continued for 6 weeks. Some mice at 8 weeks of age were fed mouse chow with 4% fat and no added cholesterol for 6 weeks. Lesions were quantified by the en face technique (see Supplementary Information and references therein).

Superoxide and aconitase assays. Aortae from mice treated with or without doxycycline for 8 days were perfused then placed at 4°C and the adventitia removed. Superoxide was detected using a modification of a previously described protocol (see Supplementary Information and references therein). The arch was divided into two equal rings and the thoracic aorta into four–six rings of about 4 mm in length. Each was cut longitudinally and placed in aerated Krebs's buffer (118.3 mM NaCl, 4.7 mM KCl, 2.5 mM CaCl₂, 1.2 mM MgSO₄, 1.2 mM KH₂PO₄, 25 mM NaHCO₃, 11 mM glucose) at 37°C for 30 min. Aortic sheets were then transferred to scintillation vials with 1 ml buffer ± 5 μM lucigenin (bis-*N*-methylacridinium nitrate) in the presence or absence of 150 U ml⁻¹ PEG-SOD. Samples were dark-adapted for 15 min at 37°C then chemiluminescence was measured using a scintillation counter. Aconitase activity was assayed using reagents in kit form (Oxis). Aortae from mice treated with or without doxycycline for 8 days were cleaned and homogenized in ice-cold 0.2 mM sodium citrate in 50 mM Tris-HCl (pH 7.4) for about 30 s, then consumption of NADP⁺ was quantified by monitoring absorbance at 340 nm.

Proteomic analysis. Chow-fed *SM22-TRE-Ucp1 Apoe*^{-/-} mice were treated with or without doxycycline for 8 days. Aortae were perfused with phosphate-buffered saline, dissected, rinsed and the adventitia removed. Pooled aortae were homogenized in a Potter-Elvehjem homogenizer with 1.0 ml of lysis buffer (30 mM Tris-HCl, pH 7.8, 7 M urea, 2 M thiourea, 4% CHAPS, 1 mM NaF, 1 mM Na₃VO₄, and complete protease inhibitor cocktail) on ice for about 1 min. Samples were analysed by fluorescent two-dimensional difference gel electrophoresis and mass spectrometry as described (see Supplementary Information and references therein). Additional information is provided in Supplementary Methods.

Received 22 December 2004; accepted 1 March 2005.

- Allayee, H., Gharalpour, A. & Lusis, A. J. Using mice to dissect genetic factors in atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1501–1509 (2003).
- Heuper, W. C. General reviews. *Arch. Pathol.* **38**, 162–181 (1944).
- Santerre, R. F., Nicolosi, R. J. & Smith, S. C. Respiratory control in preatherosclerotic susceptible and resistant pigeon aortas. *Exp. Mol. Pathol.* **20**, 397–406 (1974).
- Dröge, W. Free radicals in the physiological control of cell function. *Physiol. Rev.* **82**, 47–95 (2002).
- Nohl, H. Generation of superoxide radicals as byproduct of cellular respiration. *Ann. Biol. Clin. (Paris)* **52**, 199–204 (1994).
- Skulachev, V. P. Uncoupling: New approaches to an old problem of bioenergetics. *Biochim. Biophys. Acta* **1363**, 100–124 (1998).
- Echtay, K. S. *et al.* Superoxide activates mitochondrial uncoupling proteins. *Nature* **415**, 96–99 (2002).

- Nishikawa, T. *et al.* Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. *Nature* **404**, 787–790 (2000).
- Blanc, J. *et al.* Protective role of uncoupling protein 2 in atherosclerosis. *Circulation* **107**, 388–390 (2003).
- Pettersson, G. Effect of dinitrophenol and anoxia on isometric tension in rabbit colon smooth muscle. *Acta Pharmacol. Toxicol. (Copenh.)* **57**, 184–189 (1985).
- Moessler, H. *et al.* The SM 22 promoter directs tissue-specific expression in arterial but not in venous or visceral smooth muscle cells in transgenic mice. *Development* **122**, 2415–2425 (1996).
- Manning, M. W., Cassis, L. A. & Daugherty, A. Differential effects of doxycycline, a broad-spectrum matrix metalloproteinase inhibitor, on angiotensin II-induced atherosclerosis and abdominal aortic aneurysms. *Arterioscler. Thromb. Vasc. Biol.* **23**, 483–488 (2003).
- Heaton, J. M. The distribution of brown adipose tissue in the human. *J. Anat.* **112**, 35–39 (1972).
- Szekely, M., Kellermayer, M., Cholnoky, G. & Sumegi, I. Thermoregulatory heat production by periaortic brown adipose tissue in the non-cold-acclimatized rat. *Experientia* **26**, 1314–1315 (1970).
- Kanski, J., Behring, A., Pelling, J. & Schöneich, C. Proteomic identification of 3-nitrotyrosine-containing rat cardiac proteins: effect of biological aging. *Am. J. Physiol. Heart Circ. Physiol.* **288**, H371–H381 (2005).
- Su, B. *et al.* Redox regulation of vascular smooth muscle cell differentiation. *Circ. Res.* **89**, 39–46 (2001).
- Zai, A., Rudd, M. A., Scribner, A. W. & Loscalzo, J. Cell-surface protein disulfide isomerase catalyzes transnitrosation and regulates intracellular transfer of nitric oxide. *J. Clin. Invest.* **103**, 393–399 (1999).
- Tamarit, J., Cabisco, E. & Ros, J. Identification of the major oxidatively damaged proteins in *Escherichia coli* cells exposed to oxidative stress. *J. Biol. Chem.* **273**, 3027–3032 (1998).
- Gardner, P. R. Aconitase: Sensitive target and measure of superoxide. *Methods Enzymol.* **349**, 9–23 (2002).
- Watts, H. in *Evolution of the Atherosclerotic Plaque* (ed. Jones, R. J.) 117 (Univ. Chicago, Chicago, 1963).
- Chen, J. *et al.* Hypertension does not account for the accelerated atherosclerosis and development of aneurysms in male apolipoprotein E/endothelial nitric oxide synthase double knockout mice. *Circulation* **104**, 2391–2394 (2001).
- Knowles, J. W. *et al.* Enhanced atherosclerosis and kidney dysfunction in eNOS^{-/-} apoE^{-/-} mice are ameliorated by enalapril treatment. *J. Clin. Invest.* **105**, 451–458 (2000).
- Levin, M. *et al.* Mapping of ATP, glucose, glycogen and lactate concentrations within the arterial wall. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1801–1807 (2003).
- Jennings, R. B., Kaltenbach, J. P. & Sommers, H. M. Mitochondrial metabolism in ischemic injury. *Arch. Pathol.* **84**, 15–19 (1967).
- Smith, E. B. The effects of age and of early atheromata on the intimal lipids in men. *Biochem. J.* **84**, 49P (1962).
- Smith, E. B. Lipids carried by S_f 0–12 lipoprotein in normal and hypercholesterolaemic serum. *Lancet* **2**, 530–534 (1962).
- Klein, P. D. & Johnson, R. M. Phosphorus metabolism in unsaturated fatty acid-deficient rats. *J. Biol. Chem.* **211**, 103–110 (1954).
- Hayashida, T. & Portman, O. W. Swelling of liver mitochondria from rats fed diets deficient in essential fatty acids. *Proc. Soc. Exp. Biol. Med.* **103**, 656–659 (1960).
- Cornwell, D. G. & Panganamala, R. V. Atherosclerosis: an intracellular deficiency in essential fatty acids. *Prog. Lipid Res.* **20**, 365–376 (1981).
- Sjodin, B., Hellsten Westing, Y. & Apple, F. S. Biochemical mechanisms for oxygen free radical formation during exercise. *Sports Med.* **10**, 236–254 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported by grants from the National Institutes of Health, Clinical Nutrition Research Unit, Diabetes Research and Training Center, an American Diabetes Association Mentor-Based Postdoctoral Fellowship, and by institutional resources provided by Washington University and the Siteman Cancer Center to the Proteomics Center at Washington University.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.F.S. (csemenko@im.wustl.edu).

Polo kinase links the stress pathway to cell cycle control and tip growth in fission yeast

Janni Petersen^{1,2} & Iain M. Hagan¹

Stress-activated mitogen-activated protein kinase cascades instigate a range of changes to enable eukaryotic cells to cope with particular insults. In *Schizosaccharomyces pombe* these responses include the transcription of specific gene sets and inhibition of entry into mitosis^{1,2}. The *S. pombe* stress response pathway (SRP) also promotes commitment to mitosis in unperturbed cell cycles to allow cells to match their rate of division with nutrient availability^{1,3}. The nature of this SRP function in cell cycle control is unknown. Entry into mitosis is controlled by mitosis-promoting factor (MPF; Cdc2/cyclin B) activity. Inhibitory phosphorylation of Cdc2 by Wee1 kinase inactivates MPF until Cdc25 removes this phosphate to promote mitosis⁴. The balance between Wee1 and Cdc25 activities is influenced by the recruitment of polo kinase (Plo1) to the spindle pole body (SPB)⁵. The SPB component Cut12 mediates this recruitment^{5,6}. Hyper-activating mutations in either *cut12* or *plo1* enable Cdc25-defective cells to enter mitosis^{5,7}. The hyperactive *cut12.s11* mutation suppresses *cdc25.22*, as it promotes recruitment of active Plo1 to interphase SPBs^{6,7}. Here we show that the SRP promotes phosphorylation of Plo1 on Ser 402. In unperturbed cell cycles, SRP-mediated phosphorylation of Ser 402 promotes Plo1 recruitment to SPBs and thus commitment to mitosis. Ser 402 phosphorylation also ensures efficient reinitiation of cell tip growth and cell division during recovery from particular stresses. Thus, phosphorylation of Plo1 Ser 402 not only enables SRP signalling to modulate the timing of mitotic commitment in response to nutrient status in unperturbed cycles, but also promotes the return to normal cell cycle control after stress.

Ser 10 of histone H3 is phosphorylated by several cell cycle/stress protein kinases^{8,9}. We therefore scanned the *S. pombe* genome for similar sequences and found the strongest match around Ser 402 of the fission yeast polo kinase Plo1 (Fig. 1a). Confirmation that this site is phosphorylated *in vivo* was obtained by using phospho-specific antibodies (Fig. 1b and Supplementary Fig. S1a, b). These antibodies gave a weak signal in steady-state asynchronous cell cultures, which increased markedly upon shift from 25 °C to 37 °C (Fig. 1c and Supplementary Fig. S1c), indicating that Ser 402 phosphorylation might be mediated by the heat-activated SRP. Blocking SRP signalling by deleting the mitogen-activated protein (MAP) kinase Sty1/Spc1 (*sty1.Δ*) or its upstream MAP kinase kinase Wis1 (*wis1.Δ*)¹ abolished phosphorylation of Ser 402, whereas boosting signalling through activating mutations in *wis1* (*wis1.DD*) or the deletion of *atf1*⁺ (ref. 1) enhanced both basal and heat-induced phosphorylation of Ser 402 (Fig. 1c and Supplementary Fig. S1d). This reciprocal relationship between SRP signalling and Ser 402 phosphorylation was mirrored by functional analyses that combined these mutations with a temperature-sensitive *plo1.ts2* mutation⁵. The severity of the *plo1.ts2* mutation was decreased by constitutive SRP signalling (*wis1.DD*) and enhanced by blocking SRP (*sty1.Δ*)

(Fig. 1d). SRP signalling therefore has a positive effect on Plo1 function.

Monitoring cells that had been synchronized with respect to cell cycle progression indicated that Ser 402 was phosphorylated from entry into mitosis until mitotic exit (Fig. 2a–c). Because both Plo1 and SRP signalling modulate mitotic commitment we asked whether mutation of Ser 402 to glutamic acid, to mimic phosphorylation, could compensate for the delay in mitotic commitment arising from the loss of SRP signalling (*sty1.Δ*)³. We did this by monitoring tip growth. Because tip growth continues irrespective of any perturbation of cell cycle progression, delaying mitotic commitment increases length at division, and accelerating it produces shorter cells^{10,11}. The average length of cells in a population therefore gives an accurate measure of the rate at which that particular strain commits to mitosis, and the standard deviation gives a measure of the fidelity with which cells take this decision¹⁰. Because basal SRP activity depends on nutrient status³, measurements were made in rich and minimal media. In rich medium a considerable portion of the delay in mitotic commitment arising from the deletion of *sty1*⁺/*spc1*⁺ was compensated for by mutation of Ser 402 to glutamic acid. A similar trend was apparent in minimal medium (Fig. 2d, Table 1). Therefore, in an unperturbed cell division cycle, a basal level of SRP signalling³ promotes cyclical phosphorylation of Plo1 Ser 402 to regulate the timing of mitotic commitment.

Because Cut12 also modulates Plo1 function to control mitotic commitment⁵ we asked whether SRP signalling to Plo1 worked in a similar way to Cut12. There is a direct correlation between the ability of *cut12.s11* to suppress *cdc25.22* and its ability to recruit Plo1 to the interphase SPB (Plo1 does not associate with the interphase SPB of wild-type cells)⁶. This enhanced recruitment is directly mirrored by an increase in the ability to stain interphase SPBs with the antibody MPM2, which gives an *in vivo* readout of Plo1 activity on the SPB because the appearance of the phospho-epitope it recognizes on mitotic SPBs depends on Plo1 function. For both MPM2 and Plo1 staining the interphase recruitment/activity can be seen as an increase in the proportion of cells in which a single dot is seen (interphase and early mitosis) over those mitotic cells with two strongly stained dots⁵.

As with *cut12.s11*, boosting SRP signalling with *wis1.DD* suppressed *cdc25.22* (Fig. 3a, c). This suppression depended on the phosphorylation of Ser 402 because it was not seen in a *plo1.S402A* background (Fig. 3b, c). *wis1.DD* also mirrored *cut12.s11* in doubling the ratio of single dots to double dots of Plo1 and MPM2 staining of SPBs (Fig. 3d, e). Again, the increased affinity of Plo1 for SPBs arising from enhanced SRP signalling depended on the phosphorylation status of Ser 402 because it was not seen in *wis1.DD plo1.S402A* (Fig. 3d, e). Strikingly, simply mutating Ser 402 to glutamic acid in an otherwise wild-type cell mimicked *wis1.DD* and *cut12.s11* in promoting Plo1 association with interphase SPBs that became MPM2 reactive (Fig. 3d, e). Thus, SRP-mediated phosphorylation of Ser 402

¹Cancer Research UK Cell Division Group, Paterson Institute for Cancer Research, Wilmslow Road, Manchester M20 4BX, UK. ²The Rockefeller University, Box 5, Room 801, WRB, 1230 York Avenue, New York, New York 10021, USA.

promotes its affinity for the SPB and is required for the suppression of *cdc25.22* by enhanced SRP signalling.

If Ser 402 phosphorylation promotes Plo1 association with the SPB simply to activate Plo1, then Ser 402 status should not reduce the potency of independently activated Plo1. However, when Ser 402 was changed to alanine simultaneously with the introduction of two conserved mutations that have previously been shown to lead to the constitutive activation of polo kinases in diverse systems¹² including fission yeast⁵, the ability of the *plo1.con* mutant to suppress the *cdc25.22* mitotic entry defect was reduced⁵ (Fig. 3f). Finally we asked whether *cut12.s11* promotes mitosis in *cdc25.22* by obviating the need for Ser 402 phosphorylation. Monitoring mitotic commitment and MPM2 and Plo1 reactivity of SPBs in *cdc25.22 cut12.s11 plo1.S402X* mutants showed that the impact of *cut12.s11* was reduced by mutating Ser 402 to alanine but not to glutamic acid (Fig. 3g–i). This indicated that *cut12.s11* does not work by simply overriding the need for SRP signalling to Plo1S402.

Our initial experiments had shown that Ser 402 phosphorylation was greatly enhanced by switching cultures from 25 °C to 37 °C (Fig. 1c and Supplementary Fig. S1c). Because temperature shifts transiently arrest the commitment to mitosis¹⁰ it was possible that the stimulation of Ser 402 phosphorylation that follows a heat shift is

simply a secondary consequence of the synchronization of mitotic entry during recovery. We therefore shifted small early G2 cells to 37 °C and found phosphorylation long before these cells entered mitosis (Supplementary Fig. S2a). Furthermore, phosphorylation was still seen at 37 °C in *cdc25.22* cells that cannot enter mitosis at the higher temperature (Supplementary Fig. S2b). Phosphorylation of Ser 402 in response to heat shift is therefore not simply a consequence of the synchronization of cell division; rather, it is an independent phosphorylation induced by stress alone.

The mechanism by which Sty1/Spc1 is activated in response to heat stress differs from the way in which it is activated by osmotic and oxidative stresses. Both latter stresses lead to a strong stimulation of the MAP kinase kinase Wis1. In contrast, Wis1 stimulation in response to heat stress is moderate and the activation of Sty1/Spc1 arises from sequestration of the inactivating phosphatase, Pyp1, into an insoluble fraction that cannot access and dephosphorylate Sty1/Spc1 (refs 13, 14). Phosphorylation on Plo1 Ser 402 provides another example of a distinction between heat and osmotic/oxidative stresses, because it was not seen after oxidative or osmotic stress (Supplementary Fig. S1e).

The efficiency with which *plo1.S402A* mutants divided during the recovery from the stress of a shift from 25 °C to 37 °C was markedly

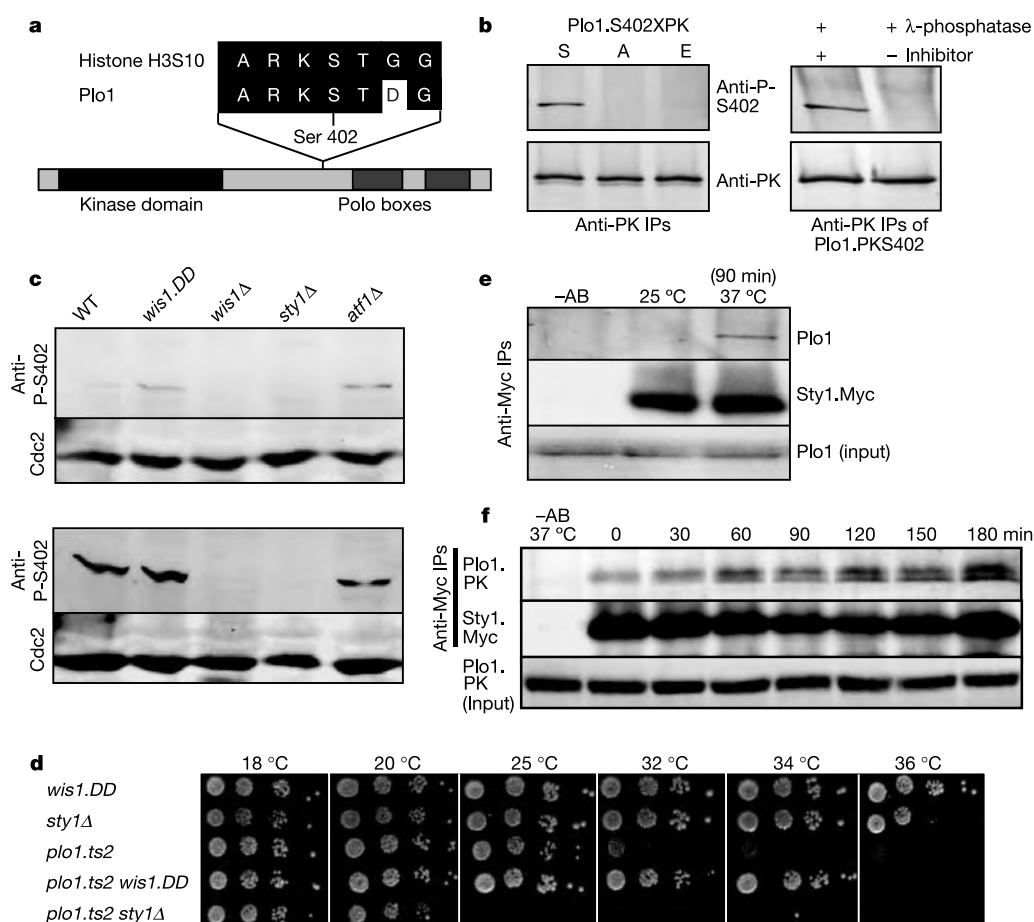


Figure 1 | SRP-dependent phosphorylation of Plo1 on Ser 402 influences Plo1 function. **a**, Residues 7–13 of histone H3 aligned with residues 399–405 of Plo1. Although the ST motif is a consensus for a dimerization site, we see no evidence for it (refs 26, 27; see Supplementary Fig. S4). **b**, Left: immunoprecipitates (IPs) of PK-tagged *plo1.S402X* mutants probed with antibodies against phospho-Ser 402. Right: treatment of PK-tagged and precipitated Plo1 with λ phosphatase with or without inhibitor. S, Plo1; A, Plo1.S402A; E, Plo1.S402E. **c**, Blots of the indicated strains with anti-phospho-Ser 402 and anti-cdc2 antibodies. Cells were

shifted from 25 °C (top) to 37 °C for 1 h (bottom). WT, wild type. **d**, Spot tests on YES medium. **e**, Blots of cell extracts (input) or anti-Myc (9E10) immunoprecipitates of *spc1.myc* cells or probed with 9E10 or Plo1 under the indicated conditions (for kinetics see Supplementary Fig. S2d, e). **f**, Blots of cell extracts (input) or anti-Myc (9E10) immunoprecipitates of *spc1.myc p81NpkPlo1* cells probed with anti-PK (mAb336) or Plo1 at the indicated times after being shifted from 25 °C to 37 °C. In **e** and **f**, –AB indicates no antibody in the immunoprecipitate.

reduced (Fig. 4a) and very few mutant cells entered mitosis within one generation time (2.2 h) at this temperature (Fig. 4b). We assume that this poor return of *plb1.S402A* reflects deficiencies in MPF activation. We also noted an interesting paradox. Despite the delay in mitotic entry in the *plb1.S402A* cells, the long mononucleated cells that should arise from this G2 arrest actually took longer to appear in *plb1.S402A* than *plb1⁺* cultures (Fig. 4c). The delay in entering mitosis meant that the opposite should have occurred; more cells should have accumulated sooner in the *plb1.S402A* mutant. This discrepancy indicated that mutation of Ser 402 might have affected tip growth after heat shock.

We therefore used fluorescent lectin to stain the glycoproteins in the cell walls and study tip growth¹⁵. We used a *cdc25.22* background for this analysis because it would hold cells in an interphase state of active tip growth, thus avoiding the complications arising from the natural cessation of tip elongation that accompanies mitosis¹¹. Cells were centrifuged out of a 25 °C culture, stained with fluorescein isothiocyanate (FITC)-lectin, reintroduced into medium at 37 °C and observed 5 h later. Insertion of fresh, unstained glycoprotein at cell tips between the immersion in FITC-lectin and observation displaces fluorescent lectin–glycoprotein complexes from tips and

Table 1 | Length of dividing cells at 25 °C

Strain	Cell length (μm)	
	Rich medium	Minimal medium
<i>plb1⁺</i>	13.1 ± 0.8	13.1 ± 0.7
<i>plb1.S402A</i>	13.3 ± 0.7	13.8 ± 1.5
<i>plb1.S402E</i>	13.1 ± 0.6	12.3 ± 1.9
<i>sty1::ura4⁺</i>	17.4 ± 1.3	21.6 ± 2.0
<i>sty1::ura4⁺ plb1.S402A</i>	17.1 ± 1.1	23.1 ± 2.8
<i>sty1::ura4⁺ plb1.S402E</i>	14.6 ± 0.9	17.5 ± 1.6

Cell lengths are shown as means ± s.d.

they become dark. If cells do not grow, there is no insertion of new material at the tips and they remain fluorescent¹⁵. At 4 h after the temperature shift, *plb1⁺* and *plb1.S402E* cells had resumed tip growth, whereas 78% of *plb1.S402A* cells had not because they still had fluorescent tips (Fig. 4e). Forward-scatter fluorescence-activated cell sorting (FACS) analysis supported this view: at least 60% of the *plb1.S402A* population had failed to show any increase in size 5 h after shift (Fig. 4d). Because F-actin patches associate with actively

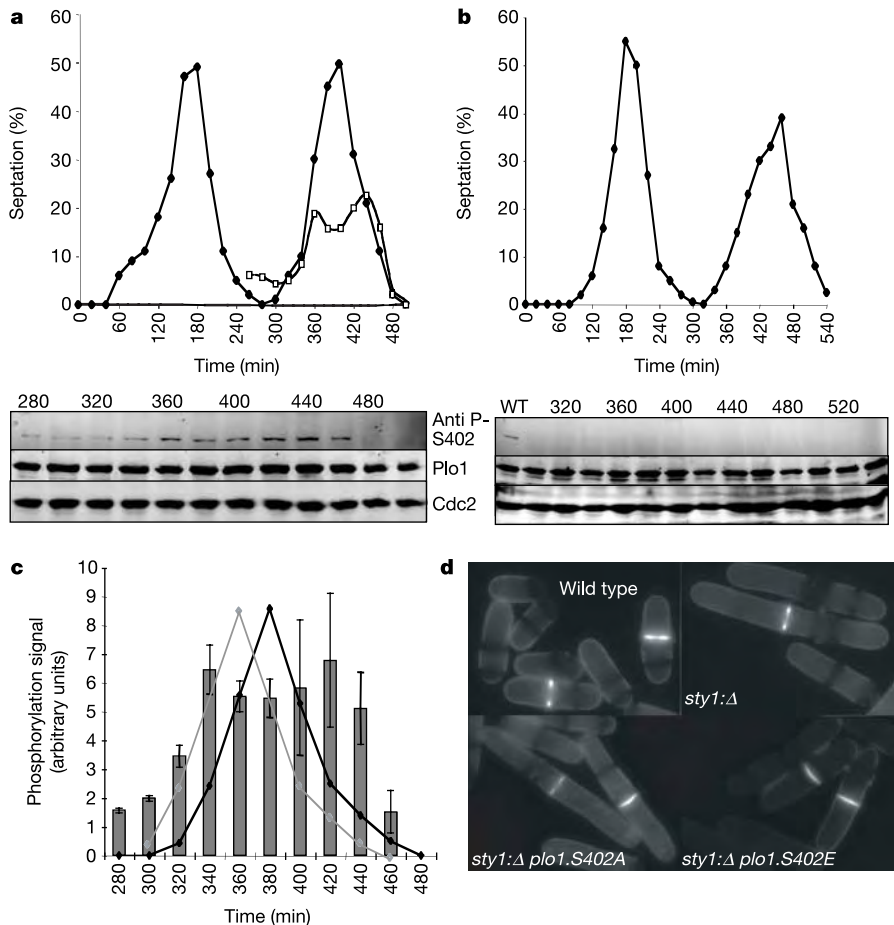


Figure 2 | Ser 402 is phosphorylated in a manner dependent on cell cycle position and SRP, and can be mutated to compensate for the *sty1*Δ mitotic commitment defect. **a–c**, Small G2 cells were isolated by centrifugal elutriation at zero time from wild-type (**a**, **c**) or *sty1::ura4⁺* (**b**) cultures at 25 °C in supplemented EMM2 medium at zero time for western blotting with anti-phospho-Ser 402 antibodies. The first cycle was ignored in case the centrifugation induced a stress response. In **a**, filled circles show septation and open circles the phospho-Ser 402 level. In **b**, WT indicates a heat-shifted wild-type control. The Plo1 loading control is 5-bromo-4-chloroindol-3-yl phosphate (BCIP) alkaline phosphatase re-probing of the anti-phospho-

Ser 402 BCIP alkaline phosphatase blot. Plo1 and Cdc2 levels do not alter with cell cycle progression^{6,28} and so mirror one another. **c**, The ratio of the phospho-Ser 402:cdc2 signal intensity from three independent cultures was normalized between experiments by equating the peaks and baselines and allocating points between them on a linear scale. The results (bars; means ± s.d.) are plotted under an average of the septation indices (black lines) from all three experiments. In these conditions mitosis (grey lines) precedes septation by 20 min (refs 11, 23). **d**, Calcofluor staining of cells in mid-exponential phase in EMM2 medium at 25 °C.

growing but not inactive tips¹⁶, we localized F-actin with rhodamine-phalloidin. Patches were found at the growing tips of *plp1*⁺ and *plp1.S402E* but were rarely seen at the non-growing tips of *plp1.S402A* cells (Fig. 4f).

Because centrifugation activates the SRP¹³ we asked whether the effects on tip growth in Fig. 4d–f were due to heat shock alone or the combination of heat and centrifugation. Application of either stress alone displaced actin from tips and stimulated Ser 402 phosphorylation (Fig. 4g, h and Supplementary Fig. S3). In both cases the return of F-actin to the cell tips was delayed by the *plp1.S402A* mutation. Repeating the combined centrifugation and heat-shift regimen of Fig. 4d–f led to an extended and more pronounced deficiency in the return of F-actin to the tips than either treatment alone (Fig. 4i). We

draw four conclusions from these observations: first, cell tip growth is blocked after either a heat shock or centrifugation; second, the kinetics of Plp1 phosphorylation is dependent on the particular stress experienced; third, the effect of the two stresses on tip growth is additive; and fourth, SRP-dependent phosphorylation of Ser 402 promotes the resumption of tip growth after perturbation by stress. Stress-induced inhibition of tip extension has until now remained cryptic owing to the efficiency with which Ser 402 phosphorylation stimulates the reinitiation of growth after stress. All interphase-arrested *cdc15* mutants continue tip extension after heat shift, irrespective of their execution point¹⁷. This indicates that tip growth can recover from heat shock at any point in the cycle. We therefore anticipate that the mechanism by which Ser 402 phosphorylation

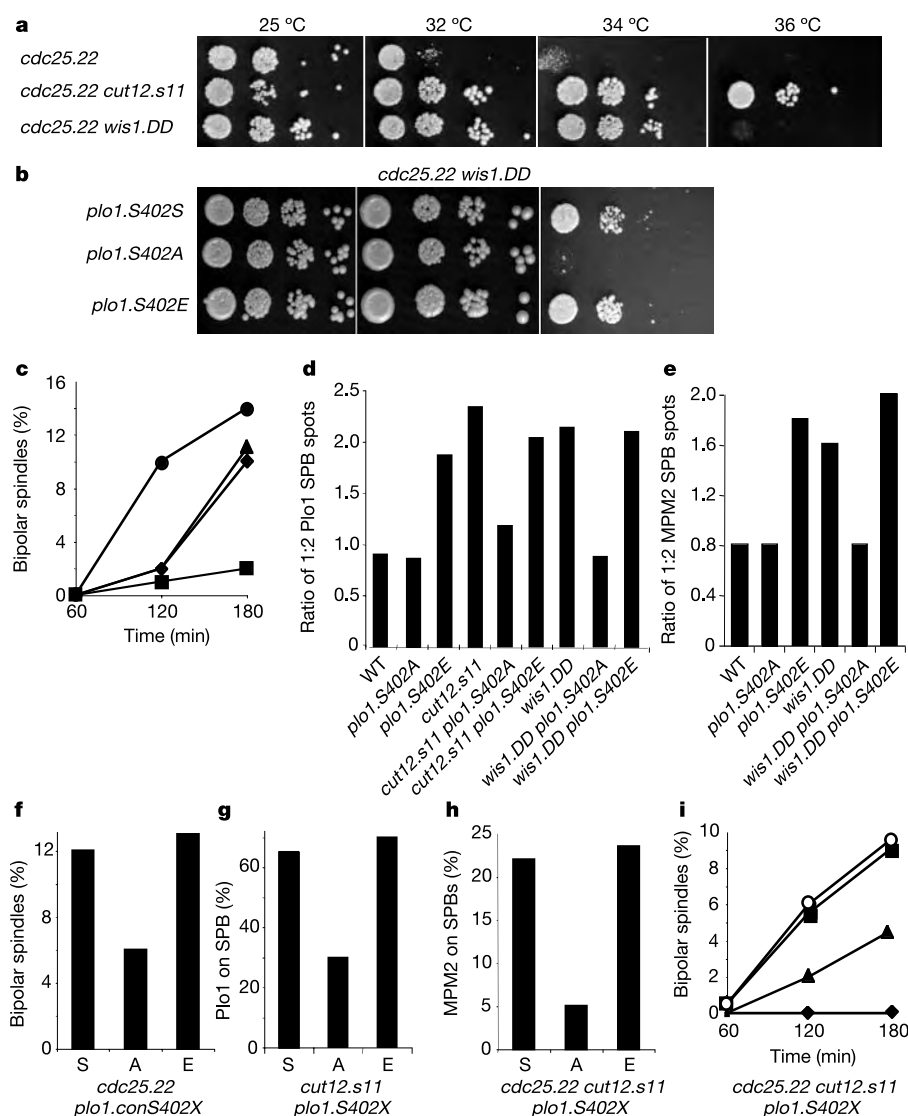


Figure 3 | Phosphorylation of Plp1 on Ser 402 promotes its association with the SPB and influences its ability to control commitment to mitosis.

a, b, Serial dilutions (from left to right) of the indicated strains were spotted onto YES (rich medium) agar plates at the indicated temperatures. The 25 °C plate is the permissive condition that enables normal growth, but the increasing temperature compromises Cdc25 function. **c–i**, Strains were grown in YES (**c–f**) or EMM2 (**g–i**) and processed for staining with HN184 polyclonal Plp1 or monoclonal MPM2 or TAT1 antibodies; at least 200 cells were scored according to the criteria indicated on the axes. **c**, Cells shifted from early exponential phase at 25 °C to 36 °C for the indicated durations. Diamonds, *cdc25.22 wis1.DD plp1*⁺; squares, *cdc25.22 wis1.DD*

plp1.S402A; triangles, *cdc25.22 wis1.DD plp1.S402E*; circles, *cdc25.22 cut12.s11*. **d, e**, Cells grown at 25 °C. WT, wild type. **f**, Cells were grown at 25 °C in EMM2 medium plus thiamine, washed three times and resuspended in EMM2; 19 h later, cells were shifted to 34 °C and fixed after a further 3.5 h. **g**, Cells were grown at 25 °C; 5 h before fixation, hydroxyurea was added to a concentration of 10 mM. **h, i**, Cells were grown to mid-exponential phase at 25 °C in EMM2 and stained with MPM2 (**h**) or shifted to 36 °C and spindles scored (**i**). In **f–h**, S, *plp1*⁺; A, *plp1.S402A*; E, *plp1.S402E*. Diamonds, *cdc25.22*; squares, *cdc25.22 cut12.s11*; triangles, *cdc25.22 cut12.s11 plp1.S402A*; circles, *cdc25.22 cut12.s11 plp1.S402E*.

promotes the reinitiation of tip extension is distinct from its role in promoting mitosis.

Comparing the kinetics of Ser 402 phosphorylation after a shift from 25 °C to 37 °C with those with which the activating phosphates appeared on Thr 171 and Tyr 173 of the MAP kinase Sty1/Spc1 supports a role for Ser 402 phosphorylation in recovery rather than the initial stress response. These phosphorylations had completely receded by the time that phosphorylation on Ser 402 of Plo1 peaked at 90 min (Supplementary Fig. S2c). It is therefore unlikely that Sty1/Spc1 directly phosphorylates Ser 402. However, the recovery kinase that does so is likely to associate with Sty1/Spc1 and Plo1 in a common complex because the ability to detect Plo1 in Sty1/Spc1

immunoprecipitates was enhanced after heat shock (Fig. 1e, f and Supplementary Fig. S2d, e). The kinetics of this association directly mirrored the kinetics with which Ser 402 became phosphorylated during the heat shift (Fig. 1f and Supplementary Figs S1c and S2).

The two functions that require the phosphorylation of Ser 402, cell cycle commitment and recovery from stress, are summarized in the diagram in Supplementary Fig. S5. The involvement of polo kinase in recovery from stress is strikingly reminiscent of its role in adaptation to DNA-checkpoint-mediated cell cycle arrest^{18–20}. Furthermore, the dependence of SRP signalling flux on nutrient status^{3,21} indicates that the phosphorylation of Plo1 on Ser 402 might be one of the methods by which metabolism and growth are coupled to cell cycle control in

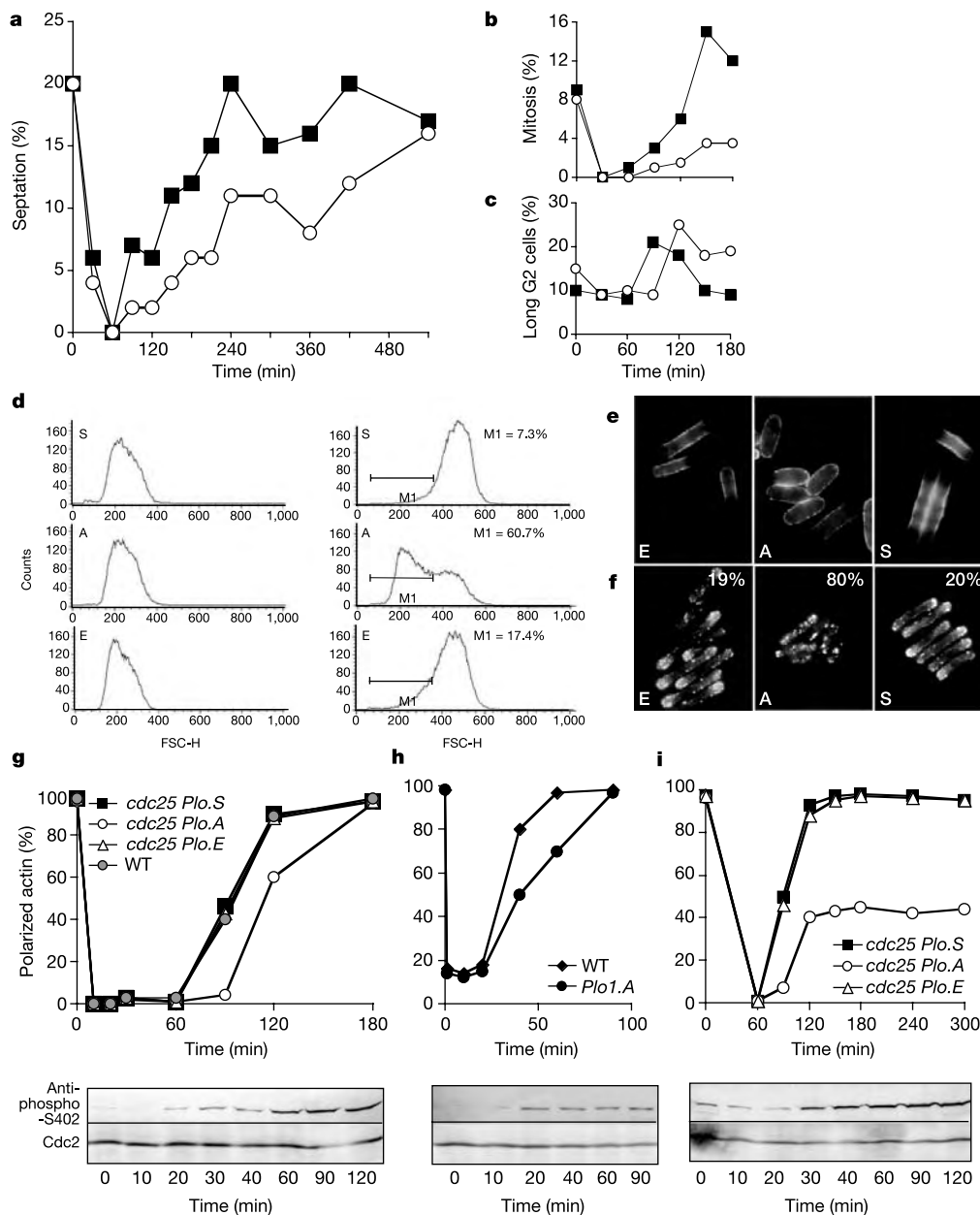


Figure 4 | SRP dependent phosphorylation of Plo1 Ser 402 promotes tip growth and cell division after stress. **a–c**, Early-exponential-phase *plo1⁺* (squares) and *plo1.S402A* (circles) cultures, grown in EMM2 medium at 25 °C, were shifted to 37 °C at zero time. Plots of septation index (**a**), mitotic index (**b**), mononucleate cells (**c**) that were of the correct size, or larger, to undergo mitosis. **d–f**, Early-exponential-phase *cdc25.22 plo1.S402X* cultures in EMM (100 mg l⁻¹) were shifted from 25 °C to 37 °C at zero time. S, *plo1⁺*; A, *plo1.S402A*; E, *plo1.S402E*. **d**, Forward-scatter FACS analysis at 0 h (left)

and 5 h (right); 20,000 counts for each strain at 25 °C; 30,000 at 37 °C. **e, f**, Cells were transiently exposed to fluorescent GS-IB4 lectin before temperature shift to 37 °C for 4 h and phalloidin staining. Staining with lectin (**e**) and F-actin (**f**) of *plo1.S402X* strains. **g–i**, Mid-exponential-phase 25 °C cultures shifted to 37 °C (**g**), centrifuged three times at 2,900 g for 2 min (**h**) or both centrifuged and shifted to 37 °C (**i**). Panels show the frequency of polarized actin and blots with anti-phospho-Ser 402 and anti-Cdc2. WT, wild type.

fission yeast. Thus, polo kinases might have a more widespread function in moderating cellular responses to changes in the environment than previously realized.

METHODS

Cell cultures and strains. Strains are listed in Supplementary Table 1. Cells were cultured in EMM2 medium supplemented with either 100 or 250 mg l⁻¹ amino acids or yeast extract medium (YES)²². Thiamine (4 µM) was added to minimal medium to repress the *nmt* promoters²³. The genomic mutation of Ser 402 to alanine or glutamic acid was generated by using the marker-switch approach that leaves a *ura4⁺* gene upstream of the *plol* open reading frame²⁴. The *ura4⁺* gene was subsequently removed by transformation in the wild-type DNA covering the region of integration by selecting transformants on FOA. We confirmed that this generated strains in which the only change from the wild type was the mutation of Ser 402 of *plol*⁺ to alanine or glutamic acid by sequencing polymerase-chain-reaction-amplified genomic DNA. These strains were then used in crosses against *plol*⁺:*ura4⁺* or *plol*⁺:*LEU2⁺* strains to ensure that daughter cells contained the appropriate *plol*.*S402X* mutation. When cells were isolated from culture by centrifugation they were subjected to 2,900 g for 2 min per spin. During lectin staining, cells were isolated by this centrifugation and subjected to two more spins as they were washed with fresh medium before reinoculation.

Molecular genetics. Ser 402 was mutated to alanine or glutamic acid in pRep81-PK-N-*plol* or in pINT6-41*plol*.con⁵. The *plol*.con-containing *NotI* fragment was subsequently integrated into the *leu1* locus of IH635.

Microscopy, FACS analysis and antibody production. Calcofluor staining of septa, rhodamine-phalloidin staining of F-actin and HN184 anti-*Plol*, TAT1 α-tubulin and MPM2 immunofluorescence procedures were as described previously^{5,6,16,22}. For lectin staining of cell walls, cells were washed in fluorescent lectin (catalogue no. I-21411; Molecular Probes; 1:50 dilution of a 0.5 mg ml⁻¹ stock solution), washed and resuspended in medium¹⁵. At least 200 cells were counted for each time point. Images and cell length measurements were obtained with a Quantix camera (Photometrics) with Metamorph software (Universal Imaging). More than 100 dividing cells per strain were measured for cell length data. Cells were fixed in 3% formaldehyde for 10 min before forward-scatter FACS analysis. Phospho-specific antibodies to a 12-amino-acid *Plol* Ser 402-phosphorylated peptide was raised and purified by Eurogentec.

Biochemistry. Total extracts were prepared by precipitation with trichloroacetic acid²⁵, denatured at 80 °C for 2.5 min and run on a 1.5-mm 7.5% SDS polyacrylamide gel. The proteins were transferred to poly(vinylidene difluoride) by wet blotting for 25 min at 100 V. Blots were probed with anti-phospho-Ser 402 (dilution 1:1,000) and secondary anti-rabbit alkaline-phosphatase-conjugated antibodies (dilution 1:10,000). *Plol* immunoprecipitation was as described previously⁵. λ phosphatase (New England Biolabs) treatment of immunoprecipitates was in accordance with the manufacturer's instructions. Immunoprecipitation of Sty1/Spcl.myc (Spcl-12myc) was performed as described in ref. 14 with magnetic Protein G beads (Dyna) to isolate the immune complexes. Modified Sty1/Spcl was recognized with PTPY antibodies against phospho-p38 MAP kinase (catalogue no. 9211S; Cell Signaling) and PY clone 4G10 anti-phosphotyrosine (catalogue no. 05-321X; Upstate).

Received 1 December 2004; accepted 4 April 2005.

- Toone, W. M. & Jones, N. In *The Molecular Biology of Schizosaccharomyces pombe* (ed. Egel, R.) 57–72 (Springer, Berlin, 2004).
- Chen, D. et al. Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell* **14**, 214–229 (2003).
- Shiozaki, K. & Russell, P. Cell-cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. *Nature* **378**, 739–743 (1995).
- Nurse, P. Universal control mechanism regulating onset of M-phase. *Nature* **344**, 503–508 (1990).
- MacIver, F. H., Tanaka, K., Robertson, A. M. & Hagan, I. M. Physical and functional interactions between polo kinase and the spindle pole component Cut12 regulate mitotic commitment in *S. pombe*. *Genes Dev.* **17**, 1507–1523 (2003).
- Mulvihill, D. P., Petersen, J., Ohkura, H., Glover, D. M. & Hagan, I. M. *Plol* kinase recruitment to the spindle pole body and its role in cell division in

- Schizosaccharomyces pombe*. *Mol. Biol. Cell* **10**, 2771–2785 (1999).
- Hudson, J. D., Feilottter, H. & Young, P. G. *stf1*: non wee mutations epistatic to *cdc25* in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **126**, 309–315 (1990).
- Soloaga, A. et al. MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *EMBO J.* **22**, 2788–2797 (2003).
- Nowak, S. J. & Corces, V. G. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet.* **20**, 214–220 (2004).
- Nurse, P. Genetic control of cell size at cell division in yeast. *Nature* **256**, 547–551 (1975).
- Mitchison, J. M. & Nurse, P. Growth in cell length in the fission yeast *Schizosaccharomyces pombe*. *J. Cell Sci.* **75**, 357–376 (1985).
- Qian, Y. W., Erikson, E. & Maller, J. L. Mitotic effects of a constitutively active mutant of the *Xenopus* polo-like kinase Plx1. *Mol. Cell. Biol.* **19**, 8625–8632 (1999).
- Shiozaki, K., Shiozaki, M. & Russell, P. Heat stress activates fission yeast Spc1/Styl MAPK by a MEKK-independent mechanism. *Mol. Biol. Cell* **9**, 1339–1349 (1998).
- Nguyen, A. N. & Shiozaki, K. Heat-shock-induced activation of stress MAP kinase is regulated by threonine- and tyrosine-specific phosphatases. *Genes Dev.* **13**, 1653–1663 (1999).
- May, J. & Mitchison, J. M. Length growth in fission yeast cells measured by two novel techniques. *Nature* **322**, 752–754 (1986).
- Marks, J. & Hyams, J. S. Localization of F-actin through the cell-division cycle of *Schizosaccharomyces pombe*. *Eur. J. Cell Biol.* **39**, 27–32 (1985).
- Nurse, P., Thuriaux, P. & Nasmyth, K. Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* **146**, 167–178 (1976).
- Toczyski, D. P., Galgoczy, D. J. & Hartwell, L. H. CDC5 and CKII control adaptation to the yeast DNA damage checkpoint. *Cell* **90**, 1097–1106 (1997).
- Yoo, H. Y., Kumagai, A., Shevchenko, A. & Dunphy, W. G. Adaptation of a DNA replication checkpoint response depends upon inactivation of claspin by the Polo-like kinase. *Cell* **117**, 575–588 (2004).
- Van Vugt, M. A., Bras, A. & Medema, R. H. Polo-like kinase-1 controls recovery from a G2 DNA damage-induced arrest in mammalian cells. *Mol. Cell* **15**, 799–811 (2004).
- Ogden, J. E. & Fantes, P. A. Isolation of a novel type of mutation in the mitotic control of *Schizosaccharomyces pombe* whose phenotypic-expression is dependent on the genetic background and nutritional environment. *Curr. Genet.* **10**, 509–514 (1986).
- Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
- Maundrell, K. Thiamine-repressible expression vectors pREP and pRIP for fission yeast. *Gene* **123**, 127–130 (1993).
- MacIver, F. H., Glover, D. M. & Hagan, I. M. A 'marker switch' approach for targeted mutagenesis of genes in *Schizosaccharomyces pombe*. *Yeast* **20**, 587–594 (2003).
- Caspari, T. et al. Characterization of *Schizosaccharomyces pombe* Hus1: a PCNA-related protein that associates with Rad1 and Rad9. *Mol. Cell. Biol.* **20**, 1254–1262 (2000).
- Elia, A. E. et al. The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. *Cell* **115**, 83–95 (2003).
- Elia, A. E., Cantley, L. C. & Yaffe, M. B. Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. *Science* **299**, 1228–1231 (2003).
- Simanis, V. & Nurse, P. The cell-cycle control gene *cdc2⁺* of fission yeast encodes a protein-kinase potentially regulated by phosphorylation. *Cell* **45**, 261–268 (1986).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Grallert for technical assistance in the initial stages of the project; K. Gull, P. Russell, H. Ohkura and N. Jones for reagents; Paul Nurse for support; and N. Jones and P. Nurse for discussions. This work was supported by Cancer Research UK and Rockefeller University.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.M.H. (ihagan@picr.man.ac.uk).

Structural and mechanistic insights into the interaction between Rho and mammalian Dia

R. Rose^{1*}, M. Weyand^{1*}, M. Lammers^{1*}, T. Ishizaki¹, M. R. Ahmadian¹ & A. Wittinghofer¹

Formins are involved in a variety of cellular processes that require the remodelling of the cytoskeleton. They contain formin homology domains FH1 and FH2, which initiate actin assembly^{1,2}. The Diaphanous-related formins form a subgroup that is characterized by an amino-terminal Rho GTPase-binding domain (GBD) and an FH3 domain, which bind somehow to the carboxy-terminal Diaphanous autoregulatory domain (DAD) to keep the protein in an inactive conformation^{3,4}. Upon binding of activated Rho proteins, the DAD is released and the ability of the formin to nucleate and elongate unbranched actin filaments is induced. Here we present the crystal structure of RhoC in complex with the regulatory N terminus of mammalian Diaphanous 1 (mDia1) containing the GBD/FH3 region, an all-helical structure with armadillo repeats. Rho uses its 'switch' regions for interacting with two subdomains of GBD/FH3. We show that the FH3 domain of mDia1 forms a stable dimer and we also identify the DAD-binding site. Although binding of Rho and DAD on the N-terminal fragment of mDia1 are mutually exclusive, their binding sites are only partially overlapping. On the basis of our results, we propose a structural model for the regulation of mDia1 by Rho and DAD.

A large number of extracellular stimuli are known to regulate the dynamic equilibrium between monomeric G-actin and filamentous F-actin. Formins regulate the nucleation and dynamics of linear actin structures. In most cases, the formin homology FH2 domain is sufficient to mediate nucleation of actin filaments *in vitro*^{5–10} and is believed to form leaky caps at the barbed ends of actin filaments, where it mediates filament elongation^{1,2,11}. The N terminus of the formin mDia is characterized by a GBD and a structurally and functionally less-well-defined FH3 domain, which is believed to mediate the subcellular localization of mDia proteins¹². The intra-molecular interaction between DAD^{3,4} and GBD/FH3 is relieved by specific members of the Rho family of GTP-binding proteins¹³. Consistent with this, active mutants of mDia1 and Rho kinase (ROCK)^{4,14} can mimic the effect of constitutively active Rho to induce actin stress fibre formation in cells¹². To elucidate the molecular and structural basis for regulation of the Diaphanous-related formin (Drf) subfamily of proteins we have solved the three-dimensional structure of the complex between RhoC and a GBD/FH3 construct from mouse Dia1 (also known as Drf1 or Diap1, referred to as mDia from now on), and investigated the mechanistic implications.

Previous pull-down and two-hybrid assays have shown that mDia1 interacts only with RhoA, B and C, and that other mDia isoforms are less specific^{4,15–17}. For a more quantitative evaluation of the Rho–mDia interaction, we used an N-terminal fragment encompassing residues 69–451 (mDia_N, Fig. 1a) and determined the kinetic and equilibrium binding constants using stopped-flow and fluorescent mant-nucleotide-loaded Rho proteins. Association rates obtained by rapidly mixing Rho proteins with increasing concentrations of

mDia_N were plotted against the mDia_N concentration to obtain k_{on} . The dissociation rate constant k_{off} was obtained by chasing an RhoA·mGppNHP–mDia_N complex with excess unlabelled RhoA·GppNHP (GppNHP is a non-hydrolysable GTP analogue) (Fig. 1b). From k_{on} ($0.47 \mu\text{M}^{-1} \text{s}^{-1}$) and k_{off} (0.003s^{-1}) we obtain an equilibrium dissociation constant K_d (k_{off}/k_{on}) of 6 nM, comparable to the 9 nM affinity of Rhotekin (another Rho effector) to

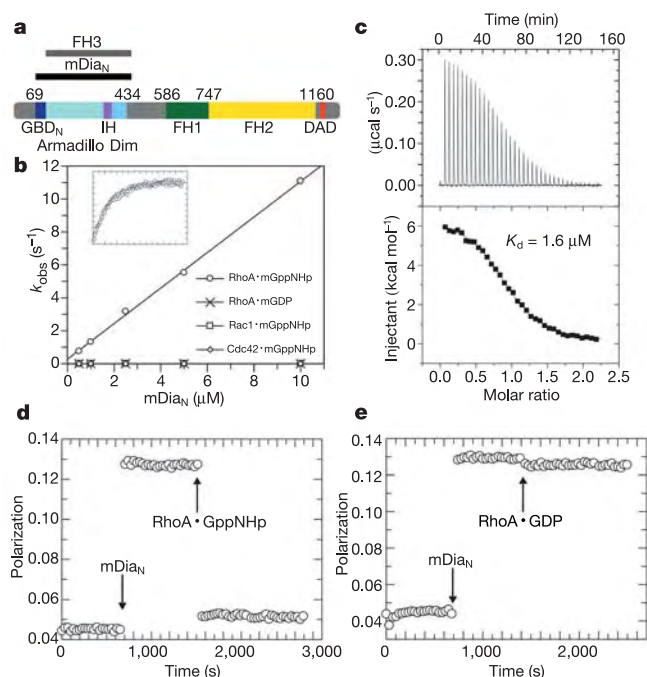


Figure 1 | Architecture and properties of mDia_N. **a**, Schematic drawing of mDia1 and the N-terminal fragment mDia_N (residues 69–451) used in the present study. Armadillo, armadillo repeat region; DAD, Diaphanous autoinhibitory domain; Dim, dimerization domain; FH1, 2, 3, formin homology domains; GBD, GTPase binding domain; IH, interdomain helix. Domain boundaries were determined by this and earlier studies^{9,10}. **b**, Specificity of mDia_N–G-protein interaction, determined by stopped-flow measurements. Observed association rate constants were plotted against mDia_N (GBD/FH3) concentration. Inset shows the dissociation of the RhoA·mGppNHP–mDia_N complex using excess unlabelled RhoA·GppNHP, plotted as relative fluorescence (arbitrary units) over time (0–1,600 s). **c**, Binding of mDia_N to DAD as determined by ITC. **d**, **e**, Fluorescence polarization of 1 μM AMCA-labelled DAD peptide, with 3 μM mDia_N and 9 μM RhoA·GppNHP (**d**) or RhoA·GDP (**e**) added at the indicated time points. All experiments were performed at least three times; typical experiments are shown.

¹Max-Planck-Institute of Molecular Physiology, Department of Structural Biology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany.

*These authors contributed equally to this work.

Table 1 | Refinement statistics

Resolution (Å)	19.8–3.0
$R_{\text{work}}/R_{\text{free}}$ (%)	21.1/28.5
Number of atoms	
Protein	8,344
GppNHp/Mg ²⁺	64/2
B-factors	
Main chain	51.1
Side chain	53.5
All atoms	52.3
r.m.s. deviations	
Bond length (Å)	0.017
Bond angles (°)	1.733

See Supplementary Table 3 for full refinement statistics.

RhoA¹⁸, and the 20 nM affinity of the Ras-binding domain of Raf-1 to Ras¹⁹. The binding of RhoA is specific for the GTP-bound conformation (Fig. 1b), confirming that the mDia_N protein behaves as a true effector. No binding was observed for Cdc42·mGppNHp or Rac1·mGppNHp, confirming the specificity of mDia1 (Fig. 1b).

In agreement with published results^{3,4,20}, mDia_N interacts with DAD. Isothermal titration calorimetry (ITC) with a 45-residue DAD peptide (1151–1196) gives a K_d of 1.6 μM (change in free energy, $\Delta G = -7.74 \text{ kcal mol}^{-1}$) with a stoichiometry of 1. Notably, the interaction of mDia_N with DAD is highly endothermic, with an enthalpy change (ΔH) of 6.45 kcal mol^{-1} (unfavourable), which is balanced by a highly favourable temperature-dependent change in entropy ($T\Delta S$) of 14.19 kcal mol^{-1} (Fig. 1c).

As both Rho and DAD can bind to the mDia_N construct, we wondered whether a ternary complex between mDia_N, RhoA·GppNHp and DAD can be formed, or whether binding is mutually exclusive. Polarization of a fluorescently labelled 22-mer DAD peptide (residues 1175–1196) was strongly increased after addition of an excess of mDia_N, owing to a decreased mobility of the higher molecular mass DAD–mDia_N complex (Fig. 1d). After addition of RhoA·GppNHp, polarization is reduced to base level,

indicating the release of the peptide from the mDia_N complex. RhoA·GDP did not cause the release of DAD, further supporting the Rho·GTP-specific release from autoinhibition of Drf proteins (Fig. 1e). During the preparation of this manuscript, a report on the mDia–DAD interaction was published²¹, with an affinity of 250 nM for a 34-mer DAD peptide. This is in agreement with our observation that DAD peptides of different lengths have different affinities for mDia1 (data not shown).

For structure determination, we used a complex of RhoC·GppNHp (henceforth called Rho) and mDia_N from mDia1 (which happened to crystallize better than the complex with RhoA). The structure was solved at 3 Å (see ref. 22 and Table 1 for refinement statistics). The crystals contained two mDia_N–Rho complexes per asymmetric unit, with very similar structures. Owing to more extensive crystal contacts of Rho in the second dimer, we limited the structural description to one of the two complexes (consisting of molecules A and B; see Fig. 2a).

mDia_N is exclusively α -helical and is comprised of the N-terminal part of the GBD (GBD_N), an armadillo repeat region (ARR) and dimerization subdomains. (Fig. 2a, b): The ARR (amino acids 136–346) is connected to the GBD_N via a loop (amino acids 124–135) that is only partially visible in the structure. Helices $\alpha 4$ – $\alpha 16$ have a repeating pattern of three helices, which is typical for armadillo repeats. The DALI server (<http://www.ebi.ac.uk/dali/>) gives the highest similarity (Z-score of 16.3) to the armadillo repeat region of β -catenin²³, although the mDia_N ARR does not have the super-helical twist of β -catenin (Fig. 2c). The last of the four armadillo repeats leads into the C-terminal dimerization domain via the interdomain helix $\alpha 17$.

The dimerization domain of mDia1 is formed by three α -helices, two of which (helices $\alpha 19$ and $\alpha 20$) form a four-helix bundle with the neighbouring molecule (Fig. 2d). There is an extensive interface between the monomers, burying a surface area of 5,422 Å² that consists mainly of hydrophobic interactions. At the edge, there are hydrophilic interactions between highly conserved residues (see

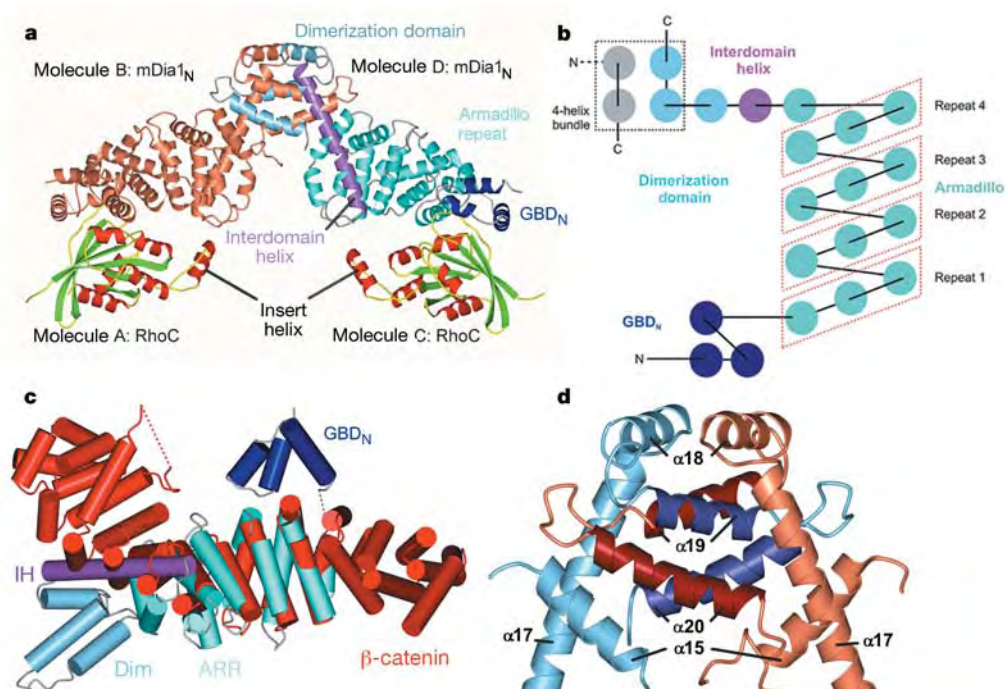


Figure 2 | Overall structure of the mDia_N–Rho complex. **a**, Ribbon representation of the RhoC·GppNHp–mDia_N dimer. Colour-coded domains from one mDia_N molecule are indicated and the second mDia_N molecule is shown in brown; RhoC·GppNHp β -strands are shown in green and helices in

red. **b**, Topology of the mDia_N structure, with colours as in **a**. **c**, Overlay of the armadillo repeats of β -catenin (red) with mDia_N. **d**, Ribbon diagram of the dimerization interface, with the helices $\alpha 19$ and $\alpha 20$ forming the four-helix bundle as shown.

Supplementary Fig. 1), such as the hydrogen bond between the invariant Ser 403^D (superscript D indicates a residue in mDia) in helix α 19 and a conserved Glu 427^D in α 20.

Confirming our structural findings, mDia_N eluted from a gel filtration column with an elution volume corresponding to a molecular mass of 160 kDa, and the complex with Rho-GppNHP did not appreciably increase the elution volume. This is consistent with the extended conformation seen from the structure. Dimerization has also been found for a fragment consisting of residues 370–548 of mDia1²¹. Owing to the high conservation of the mostly hydrophobic interface and the biochemical stability of the dimer we conclude that dimerization via the FH3 domain should also occur in the full-length protein. As formins also dimerize via an N-terminal extension of the

FH2 domain^{9,10}, we wonder about the topology of the complete protein. If monomer A interacts with monomer B via both the FH3 and FH2 domains, the full-length protein would also be a dimer. Other scenarios with higher-order oligomers are also imaginable.

Rho uses both the GBD_N and the ARR subdomains for interaction with mDia_N; this interaction has an interface of 2,981 Å² (Fig. 3a, b). Apart from Pro 36^R (superscript R indicates a residue in Rho), mDia1 covers the same surface area on Rho as the considerably smaller ROCK–RhoA interface²⁴, and largely overlaps the interface formed by the RhoA–PKN (protein kinase N) interaction site II (ref. 25), both of which are ~1,500 Å² in size. As with PKN and ROCK, mDia uses a hydrophobic patch formed by Rho residues Val 38, Phe 39 and Leu 72 to form the core interface, which is further stabilized by electrostatic

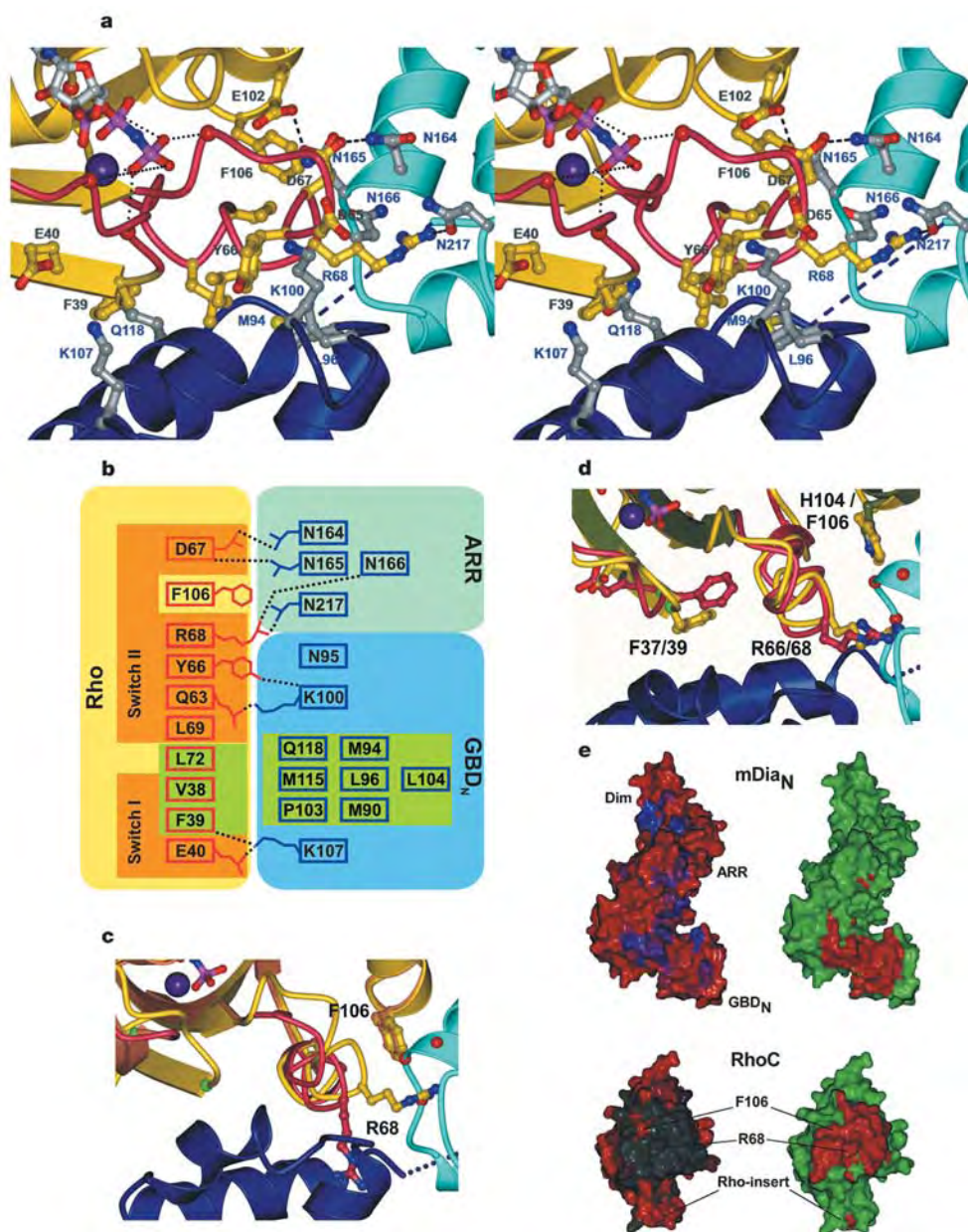


Figure 3 | The Rho-mDia_N interaction. **a**, Stereo diagram of the Rho-mDia_N interface. Colour-coding of mDia_N as in Fig. 2a, Rho in yellow, and switch regions of Rho in red. Residues involved in the interface (cutoff level 3.6 Å) are shown in ball-and-stick configuration. **b**, Schematic drawing of interacting residues in Rho and mDia. **c**, Superimposition of the Rho-mDia_N complex (yellow) with RhoA-GDP (red)²⁷ (Protein Data Bank entry 1DPF). **d**, Superimposition of the Rho-mDia_N complex with

Cdc42-GppNHP (ref. 30, Protein Data Bank entry 1AM4). **e**, Van der Waals surface representation of mDia_N (upper) and RhoC-GppNHP (lower). Left panels show residues that are 100% conserved between RhoA, B, C, Cdc42 and Rac1 (black, lower panel), or >75% (blue) or >50% (magenta) conserved between mDia1, 2, 3 and human Dia1–3 (upper panel). Residues of Rho and mDia involved in forming the Rho-mDia interface are shown in red in right panels.

interactions at the edges (Fig. 3a, b). This implies that the binding of Rho to ROCK, PKN and mDia is mutually exclusive. The Rho insert helix is close to the armadillo repeat 4 (ARM4), but does not appear to make a major contribution to the binding.

Both switch regions of Rho are involved in the interaction with mDia_N. Switch I interacts with GBD_N only, and switch II makes contact with residues in both GBD_N and the armadillo repeat subdomains. Glu 40^R, a putative Rho specificity determinant for ROCK and PKN binding²⁴, interacts with Lys 107^D and facilitates the hydrophobic interactions of Val 38^R and Phe 39^R with Leu 104^D and Met 115^D. Tyr 66^R uses a main-chain interaction to position Lys 100^D for binding to Gln 63^R. Asp 67^R, which interacts intramolecularly with Lys 98^R, contacts Asn 165^D via its main-chain oxygen. Arg 68^R is wedged into the space between the GBD_N and the first armadillo

repeat, interacting with the side chain of Asn 217^D and the main chains of Leu 163^D and Asn 166^D.

Many of the mDia1 residues in the interface are conserved between Drf proteins, but only Lys 100 and Lys 107 are completely invariant, even in *S. cerevisiae* Bni1p (see Supplementary Fig. 1). Asn 217^D binds the crucial residue Arg 68^R, and is almost completely conserved. Pro 103^D, Met 115^D and Gln 118^D form a hydrophobic patch that is complementary to that of Rho. The loop between helices α 5 and α 6 of ARM1 contains a triple Asn motif (Asn 164–166) that is highly variable (TSH in mDia2 and mDia3) and is one of the determinants of specificity (see below). Two of these Asn residues are involved in the interaction with Asp 67^R (Fig. 3a, b).

Ras-like G proteins always use one or both switch regions for binding to effectors, as they change their conformation according to

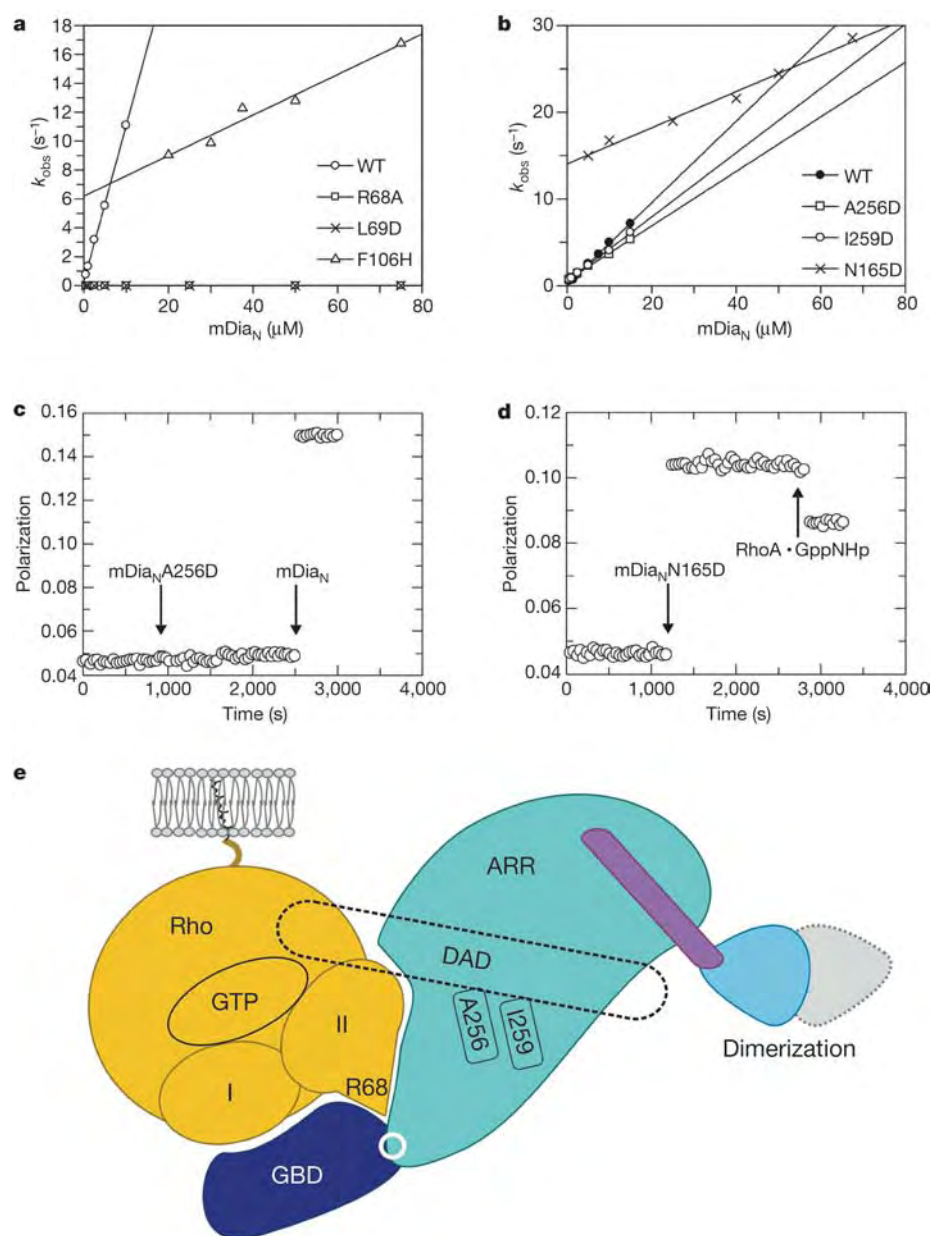


Figure 4 | Mutually exclusive binding of Rho and DAD to mDia_N.

a, b, Observed association rate constants for the binding of wild-type and mutant RhoA to wild-type mDia_N (**a**) or wild-type RhoA to different mDia_N mutants (**b**), measured as in Fig. 1. **c, d**, Fluorescence polarization assay with 1 μM AMCA-labelled DAD peptide, mDia_N mutants A256D (**c**) and N165D (**d**) and RhoA as described in Fig. 2. In **c**, wild-type mDia_N was added as a

control, to show DAD binding ability. All experiments were repeated at least twice and typical experiments are shown. **e**, Schematic diagram of the partially overlapping binding sites identified here and the proposed mechanism of release of DAD-mediated autoinhibition by Rho, as discussed in the text.

the nature of the bound nucleotide²⁶. When the structures of RhoA·GDP²⁷ (Protein Data Bank entry 1DPF) and the mDia-bound Rho·GppNHP are superimposed, structural differences in both the switch I and II regions of Rho are apparent (Fig. 3c). In switch I, interface residues Val 38^R, Phe 39^R and Glu 40^R adopt a different conformation in the absence of the γ -phosphate of GTP, thereby modifying the hydrophobic patch. This alone might account for the nucleotide-dependency of mDia binding (see below). Conformational differences are also seen for Leu 69^R and Tyr 66^R of switch II. Arg 68^R in RhoA·GDP is more disordered and also shifted several ångströms out of the Rho–mDia interface, indicating that it might be important for discrimination between Rho·GTP and Rho·GDP (Fig. 3c).

Residues found in the interface of the Rho–mDia complex are conserved between RhoA, B and C, and are somewhat different for Rac and Cdc42. To probe ligand specificity, we analysed mDia1 and RhoA mutants (Fig. 4a). R68A^R and L69D^R mutations completely abolish binding to mDia1, stressing the importance of the deeply penetrating Arg 68^R and the hydrophobic patch. When Phe 106^R was mutated to His (which is the residue found in the equivalent position in Cdc42 and Rac1), the affinity to mDia_N was greatly reduced (by more than 7,000-fold) to 44 μ M, compared with 6 nM for wild-type RhoA. This loss of affinity is primarily due to a large increase in the dissociation rate constant, from 0.003 s^{−1} to 6.2 s^{−1}. The decrease in k_{on} was much smaller (0.47 μ M^{−1}s^{−1} versus 0.14 μ M^{−1}s^{−1}). An N165D^D mutation in mDia_N also induces an increase in dissociation rate and decrease in affinity (K_d = 67 μ M; k_{on} = 0.21 μ M^{−1}s^{−1}; k_{off} = 14.03 s^{−1}) (see Supplementary Table 1).

Previous experiments have demonstrated specific interacting pairs of Rho family members and formin proteins^{4,15,17}. From the residues located in the interface with mDia1, only Glu 40^R and Phe 106^R are different between Rho isoforms A, B and C and Rac/Cdc42 (Fig. 3d). It has been shown that mutation of Phe 39^R to Ala, Val or Leu had no effect on the binding of RhoA to mDia2 in two-hybrid assays²⁸, nor did mutation of Glu 40^R. Specificity could be due to His 104 of Cdc42 and Rac1, which corresponds to Phe 106 of Rho (close to the triple Asn motif, Asn 164–166). This argues for a different set of interactions in the interface of mDia2/3 with Cdc42, as homologous residues are highly variable in Drf proteins. Supporting this, an F106H^R mutation drastically reduces Rho affinity for mDia1 (Fig. 4a and Table 1) and the H104F mutation increases the otherwise extremely low (>100 μ M) affinity of Cdc42 for mDia1 to 2 μ M (see Supplementary Table 1). A CRIB-like (Cdc42/Rac interactive binding) motif in mDia2 has been reported to be involved in binding to Cdc42 (ref. 17). However, Gln 142^D of mDia1, equivalent to His 160 from the putative mDia2 CRIB motif, is 17 Å away from Rho in our structure. As Rac and Cdc42 also seem to use the Rho binding site described here, binding of Cdc42 to the CRIB-like motif is unlikely.

The autoinhibitory domain is highly conserved in Drf proteins³ and shares a mutually exclusive binding site with Rho in the GBD/FH3 region (see also ref. 21). However, the surface of mDia1 (Fig. 3e) shows a striking discrepancy between the Rho-binding interface and the location of conserved residues, but no such discrepancy is seen between conserved Rho residues and their mDia-binding face. Apart from the dimerization domain, conserved residues in the concave side of the mDia ARR are located close to, but only partially overlapping with, the Rho-binding interface. We thus consider this patch of conserved residues—which stretches across the solvent-accessible surfaces of the C-terminal sides of helices α 5, α 8, α 11 and α 14, in a cleft between the GBD_N and the interdomain helix—to be the putative DAD binding site of mDia.

To test this, we mutated relevant residues. Both A256D^D and I259D^D mutations significantly reduced the affinity of DAD to the mDia_N construct without affecting Rho binding appreciably, as shown by ITC and stopped-flow measurements (Fig. 4b, c and Supplementary Table 1). In contrast, even rather drastic mutations

of residues located on the convex surface of the ARR, such as E264K^D and R269E^D, did not affect the interaction of DAD with mDia_N (see Supplementary Table 1). The N165D^D mutant has a greatly reduced affinity to Rho, but the binding of a DAD fragment is only slightly affected (Fig. 4b, d and Supplementary Table 1). Accordingly, as shown by fluorescent polarization (Fig. 4d), Rho cannot displace DAD from the (N165D^D) Dia mutant.

Active mDia1 can induce the formation of disorganized actin stress fibres in the absence of ROCK, but normally the balance of activity of these two proteins determines the thickness and density of induced stress fibres⁴. In contrast with dominant active RhoA(G14V), which induces bundled stress fibres throughout the cells (see Supplementary Fig. 2), the induction of stress fibres is reduced by L69D^R and R68A^R mutations. The latter mutant has a less severe phenotype and shows some stress fibres, but the Leu 69 mutation results in elongated morphology and no stress fibres. Although the reason for the differences in phenotype is not obvious at the moment, it is not due to differences in the interaction with ROCK, because both mutations equally abolish the interaction with the Rho-binding domain of ROCK (not shown).

By introducing appropriate mutations, we have identified the DAD-binding site as a highly conserved patch of residues that is only partially overlapping and can be uncoupled from the Rho-binding site. DAD presumably assumes a helical structure because it has a high helix propensity, and part of the DAD domain structure has been identified as a helix in the structure of the C-terminally elongated FH2 domain from mDia1 (ref. 9). We thus propose that binding of an elongated amphipathic DAD helix would sterically clash with the Rho-binding site identified in our structure (Fig. 4e). Although the driving force for releasing an intramolecular interaction by an intermolecular protein–protein interaction is unclear, partially overlapping binding sites for DAD and Rho should be favourable. As Rho has a bipartite binding site on mDia_N, we envision a two-step binding reaction where a low-affinity ternary complex would first be formed, in which Rho sterically interferes with DAD binding. After release of DAD, it would assume the tight binding conformation seen in our structure. As GBD_N is only weakly connected to the ARR, a conformational change during the two-step binding is feasible (Fig. 4e). This would be similar to the mechanism of nucleotide exchange on Ras-like proteins, where a guanine-nucleotide exchange factor (GEF) first weakens nucleotide affinity in a ternary complex and then releases the nucleotide to form a tight, binary GEF complex²⁶.

How release of autoinhibition activates the actin nucleation potential of the FH2 domain of mDia is unclear and will require additional structural investigations on larger mDia constructs. The structure of the mDia–Rho complex and the biochemical experiments presented here are an important step towards understanding the molecular mechanisms of Diaphanous-related formin regulation.

METHODS

Structure determination. Purification, crystallization and data collection of native crystals are described elsewhere²². Crystals belonged to the orthorhombic space group $P2_12_12_1$ and contained two mDia_N–RhoC dimers in the asymmetric unit. Out of 38 selenium atoms, 12 were found by SOLVE (<http://www.solve.lanl.gov/>) using the anomalous signal of the selenomethionine-substituted crystals compared to wild-type crystals. Selenium sites were refined and initial phases were calculated using the program SHARP (<http://www.globalphasing.com/>), yielding 17 additional sites for a total of 29.

The program O (http://xray.bmc.uu.se/~alwyn/Distribution/distrib_frame-set.html) was used to build the model into the $2F_o - F_c$ and $F_o - F_c$ maps in iterative rounds of refinement with REFMAC (<http://www.ysbl.york.ac.uk/~garib/refmac/index.html>) and CNS (<http://cns.csb.yale.edu/v1.1/>). The complete chains A–D were defined as TLS (translation, libration and screw) groups for refinement. Model bias was avoided by calculating composite simulated omit maps. The CCP4 suit was used extensively during refinement (<http://www.ccp4.ac.uk/>) (see Supplementary Table 2 for refinement statistics). The final model has a good geometry, with 99% of all residues in the allowed

regions of the Ramachandran plot as judged by the program Procheck (<http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html>). Ribbon plots were prepared using Molscript (<http://www.avatar.se/molscript/>) and rendered with Povray (<http://www.povray.org>).

DAD peptides. A DAD peptide containing residues 1175–1196 with and without an N-terminal AMCA fluorophore label were synthesized by Biosynthon. A construct including residues 1151–1196 was prepared as described for mDia_N, except that the GST fusion protein was cleaved and purified on a Superdex200 gel filtration column and concentrated using Amicon centrifugation units. The concentration of DAD fragments was determined using the method in ref. 29.

Stopped-flow measurements. All measurements were performed at 20 °C using an SX18 MV Applied Photophysics apparatus. The fluorophore was excited at 350 nm, emission recorded using a 408-nm cutoff filter. A final concentration of 100 nM of GTP-binding protein bound to the respective mant-labelled nucleotides, and increasing concentrations of effector (1–75 μM) were used for the measurements, in buffer containing 50 mM Tris/HCl pH 7.5, 300 nM NaCl, 5 mM MgCl₂ and 2 mM 2-mercaptoethanol (standard buffer). Analysis of binding curves was performed using the manufacturer's software and GraFit 5.0 (Erithacus Software Ltd.). Each value of k_{obs} was determined on the basis of at least four measurements. Typical errors for k_{on} and k_{off} determinations were $0.4 \pm 0.1 \mu\text{M}^{-1} \text{s}^{-1}$ and $3 \pm 0.08 \times 10^{-3} \text{s}^{-1}$, respectively, and somewhat higher for weak binding mutants.

Isothermal titration calorimetry (ITC) measurements. mDia_N interaction with the DAD was measured using isothermal titration calorimetry. The mDia_N construct (20 μM) in standard buffer (without MgCl₂) was thermostated in the cell at 20 °C. The DAD fragment containing residues 1151–1196 (at a concentration of 200 μM) was injected stepwise into the solution. The change in heating power was observed until equilibrium was reached before the next injection was started. The data was then analysed using software provided by the manufacturer. Typical errors for ITC measurements were $0.4 \pm 0.07 \mu\text{M}$, and somewhat higher for weak binding mutants.

Polarization assay. Polarization measurements were performed using a FluoroMax II spectrofluorimeter with a polarization filter (SPEX Instruments; excitation wavelength 353 nm, emission wavelength 442 nm) at 20 °C in standard buffer containing 5% DMSO. The AMCA-labelled DAD fragment was equilibrated at a 1 μM concentration, and a threefold excess of mDia_N (wild type or mutant) was added, rapidly mixed and fluorescence was measured. Finally, the change in polarization after addition of 9 μM RhoA-GppNHp or RhoA-GDP was recorded.

Transfection of RhoA mutants. HeLa cells were plated on a coverglass at a density of 7.5×10^4 cells per 3.5-cm dish and cultured in DMEM medium containing 10% FCS. After one day, cells were incubated for 3 h in 1 ml OPTI-MEM (Gibco) containing 1 μg of the indicated plasmids and 3.5 μl lipofectamine (Gibco). The medium was then replaced with DMEM containing 10% FCS. Sixteen hours after transfection, cells were fixed with PBS containing 3.7% formaldehyde for 15 min, then permeabilized for 10 min in PBS containing 0.1% Triton X-100. For F-actin staining, rhodamine-conjugated phalloidin (Molecular Probes) was used. Cells were analysed in 0.4-μm optical sections using a Bio-Rad MRC-1024 confocal imaging system, and built-up images were constructed.

Received 4 February; accepted 29 March 2005.

Published online 1 May 2005.

- Waller, B. J. & Alberts, A. S. The formins: active scaffolds that remodel the cytoskeleton. *Trends Cell Biol.* **13**, 435–446 (2003).
- Zigmond, S. H. Formin-induced nucleation of actin filaments. *Curr. Opin. Cell Biol.* **16**, 99–105 (2004).
- Alberts, A. S. Identification of a carboxyl-terminal diaphanous-related formin homology protein autoregulatory domain. *J. Biol. Chem.* **276**, 2824–2830 (2001).
- Watanabe, N., Kato, T., Fujita, A., Ishizaki, T. & Narumiya, S. Cooperation between mDia1 and ROCK in Rho-induced actin reorganization. *Nature Cell Biol.* **1**, 136–143 (1999).
- Pruyne, D. *et al.* Role of formins in actin assembly: nucleation and barbed-end association. *Science* **297**, 531–532 (2002).
- Sagot, I., Rodal, A. A., Moseley, J., Goode, B. L. & Pellmann, D. An actin nucleation mechanism mediated by Bni1 and profilin. *Nature Cell Biol.* **4**, 626–631 (2002).
- Kovar, D. R., Kuhn, J. R., Tichy, A. L. & Pollard, T. D. The fission yeast

cytokinesis formin Cdc12p is a barbed end actin filament capping protein gated by profilin. *J. Cell Biol.* **161**, 875–887 (2003).

- Li, F. & Higgs, H. N. The mouse Formin mDia1 is a potent actin nucleation factor regulated by autoinhibition. *Curr. Biol.* **13**, 1335–1340 (2003).
- Shimada, A. *et al.* The core FH2 domain of diaphanous-related formins is an elongated actin binding protein that inhibits polymerization. *Mol. Cell* **13**, 511–522 (2004).
- Xu, Y. *et al.* Crystal structures of a Formin Homology-2 domain reveal a tethered dimer architecture. *Cell* **116**, 711–723 (2004).
- Zigmond, S. H. *et al.* Formin leaky cap allows elongation in the presence of tight capping proteins. *Curr. Biol.* **13**, 1820–1823 (2003).
- Kato, T. *et al.* Localization of a mammalian homolog of diaphanous, mDia1, to the mitotic spindle in HeLa cells. *J. Cell Sci.* **114**, 775–784 (2001).
- Etienne-Manneville, S. & Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629–635 (2002).
- Nakano, K. *et al.* Distinct actions and cooperative roles of ROCK and mDia in Rho small G protein-induced reorganization of the actin cytoskeleton in Madin-Darby canine kidney cells. *Mol. Biol. Cell* **10**, 2481–2491 (1999).
- Watanabe, N. *et al.* p140mDia, a mammalian homolog of *Drosophila* diaphanous, is a target protein for Rho small GTPase and is a ligand for profilin. *EMBO J.* **16**, 3044–3056 (1997).
- Gasman, S., Kalaidzidis, Y. & Zerial, M. RhoD regulates endosome dynamics through Diaphanous-related Formin and Src tyrosine kinase. *Nature Cell Biol.* **5**, 195–204 (2003).
- Peng, J., Wallar, B. J., Flanders, A., Swiatek, P. J. & Alberts, A. S. Disruption of the Diaphanous-related formin Drf1 gene encoding mDia1 reveals a role for Drf3 as an effector for Cdc42. *Curr. Biol.* **13**, 534–545 (2003).
- Blumenstein, L. & Ahmadian, M. R. Models of the cooperative mechanism for Rho-effector recognition: Implications for RhoA-mediated effector activation. *J. Biol. Chem.* **279**, 53419–53426 (2004).
- Herrmann, C., Horn, G., Spaargaren, M. & Wittinghofer, A. Differential interaction of the ras family GTP-binding proteins H-Ras, Rap1A, and R-Ras with the putative effector molecules Raf kinase and Ral-guanine nucleotide exchange factor. *J. Biol. Chem.* **271**, 6794–6800 (1996).
- Krebs, A., Rothkegel, M., Klar, M. & Jockusch, B. M. Characterization of functional domains of mDia1, a link between the small GTPase Rho and the actin cytoskeleton. *J. Cell Sci.* **114**, 3663–3672 (2001).
- Li, F. & Higgs, H. N. Dissecting requirements for auto-inhibition of actin nucleation by the formin, mDia1. *J. Biol. Chem.* **280**, 6986–6992 (2005).
- Rose, R., Wittinghofer, A. & Weyand, M. The purification and crystallization of mDia1 in complex with RhoC. *Acta Crystallogr. B* **61**, 225–227 (2005).
- Huber, A. H., Nelson, W. J. & Weis, W. I. Three-dimensional structure of the armadillo repeat region of β-catenin. *Cell* **90**, 871–882 (1997).
- Dvorsky, R., Blumenstein, L., Vetter, I. R. & Ahmadian, M. R. Structural insights into the interaction of ROCK1 with the switch regions of RhoA. *J. Biol. Chem.* **279**, 7098–7104 (2004).
- Maesaki, R. *et al.* The structural basis of Rho effector recognition revealed by the crystal structure of human RhoA complexed with the effector domain of PKN/PRK1. *Mol. Cell* **4**, 793–803 (1999).
- Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299–1304 (2001).
- Shimizu, T. *et al.* An open conformation of switch I revealed by the crystal structure of a Mg²⁺-free form of RhoA complexed with GDP. Implications for the GDP/GTP exchange mechanism. *J. Biol. Chem.* **275**, 18311–18317 (2000).
- Sahai, E., Alberts, A. S. & Treisman, R. RhoA effector mutants reveal distinct effector pathways for cytoskeletal reorganization, SRF activation and transformation. *EMBO J.* **17**, 1350–1361 (1998).
- Ehresmann, B., Imbault, P. & Weil, J. H. Spectrophotometric determination of protein concentration in cell extracts containing tRNA's and rRNA's. *Anal. Biochem.* **54**, 454–463 (1973).
- Rittinger, K. *et al.* Crystal structure of a small G protein in complex with the GTPase-activating protein rhoGAP. *Nature* **388**, 693–697 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Stege and D. Kuehlmann for expert technical assistance, T. Lorenz for the cloning of RhoC, I. Schlichting and W. Blankenfeldt for X-ray data collection, and the ESRF and DESY staff for support. A.W. thanks the Deutsche Forschungsgemeinschaft for financial support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.W. (Alfred.wittinghofer@mpi-dortmund.mpg.de). Coordinates have been deposited in the Protein Data Bank under the accession code 1Z2C.

Structural basis for the regulation of tubulin by vinblastine

Benoît Gigant¹, Chunguang Wang¹, Raimond B. G. Ravelli², Fanny Roussi³, Michel O. Steinmetz⁴, Patrick A. Curmi^{5†}, André Sobel⁵ & Marcel Knossow¹

Vinblastine is one of several tubulin-targeting *Vinca* alkaloids that have been responsible for many chemotherapeutic successes since their introduction in the clinic as antitumour drugs¹. In contrast with the two other classes of small tubulin-binding molecules (Taxol² and colchicine³), the binding site of vinblastine is largely unknown and the molecular mechanism of this drug has remained elusive. Here we report the X-ray structure of vinblastine bound to tubulin in a complex with the RB3 protein stathmin-like domain (RB3-SLD). Vinblastine introduces a wedge at the interface of two tubulin molecules and thus interferes with tubulin assembly. Together with electron microscopical and biochemical data, the structure explains vinblastine-induced tubulin self-association into spiral aggregates at the expense of microtubule growth⁴. It also shows that vinblastine and the amino-terminal part of RB3-SLD binding sites share a hydrophobic groove on the α -tubulin surface that is located at an intermolecular contact in microtubules. This is an attractive target for drugs designed to perturb microtubule dynamics by interfacial interference, for which tubulin seems ideally suited because of its propensity to self-associate.

The structure of vinblastine bound to the tubulin–colchicine:RB3-SLD complex ((Tc)₂R) was determined at 4.1 Å resolution. In (Tc)₂R there is one vinblastine-binding site at the interface between two tubulin heterodimers in a head-to-tail arrangement (Fig. 1a). The vinblastine-binding site was identified by inspection of difference electron-density maps calculated using diffraction data from vinblastine-soaked (Tc)₂R crystals (the highest peak, vinblastine, is 11 σ ; the first peak at a different location is 5.5 σ) (Fig. 1b). In addition, we independently determined the C12' vinblastine atom position by deducing it from the bromide position in a C12'-bromovinblastine⁵-soaked (Tc)₂R crystal. The bromide was located in difference electron density maps of the crystal after exposure to low and high doses of X-rays (Fig. 1b), taking advantage of the expected X-ray radiation sensitivity of the C–Br bond^{6,7}. This allowed us to confirm the vinblastine orientation unambiguously in its asymmetric electron density. The vinblastine site comprises residues that, in microtubules, are located towards the inner lumen and involved in longitudinal protofilament contacts⁸. It is boxed in by loop T7, helix H10 and strand S9 in the α 2 subunit and by the carboxy-terminal turn of helix H6 and loops T5 and H6–H7 in the β 1 subunit (Fig. 1c) (for the nomenclature of tubulin secondary structure elements see ref. 9 and Fig. 1c). Vinblastine buries 80% of its surface area in the complex. It is oriented so that its catharanthine and vindoline moieties each interact with both tubulin heterodimers (Fig. 1); interestingly, it interacts to the same extent with both, as the drug buries very similar

surface areas at the contacts with α 2-tubulin (400 Å²) and with β 1-tubulin (360 Å²).

There are several structural differences between (Tc)₂R and (Tc)₂R–vinblastine at the interface of the two tubulin heterodimers (Fig. 2). In particular, the side chain of Asp β 179 moves from a position close to the nucleotide to the vicinity of the drug. In addition, the T7 loop of α 2, which is not visible in (Tc)₂R, becomes ordered. Residues α 248 and α 249 in this loop contact vinblastine. Thus, these changes are directly related to vinblastine binding. There is also some movement of residue α 349, the side chain of which contacts residue β 179 but not vinblastine, so that its conformational change is an indirect consequence of vinblastine binding.

The location of vinblastine in its complex with (Tc)₂R correlates well with data collated to identify tubulin residues that interact with vinblastine. In the β -subunit the β 177– β 215 peptide has been photochemically crosslinked to vinblastine-anthranilate¹⁰. This is consistent with contacts in the structure between the drug and residues β 177, β 179 (loop T5), β 210 (helix H6) and β 214 (loop H6–H7); modelling has suggested that some of these residues contact drugs that bind to tubulin in the same region as vinblastine¹¹. In the α -subunit, Lys α 352 (strand S9) is targeted by pironetin, which competes with vinblastine for tubulin binding¹². The side chain of Lys α 352 is not visible in the electron density, but the short distance (5.5 Å) of its C α to the nearest vinblastine atom clearly indicates that it is part of the site. The (Tc)₂R structure also accounts for the effects of chemical modifications to vinblastine on its tubulin-binding activity. For example, several derivatives of vinorelbine, a vinblastine analogue, modified on the D' ring, were synthesized; both their affinities for tubulin and their biological activities were affected¹³. These molecules are changed in a region that, in (Tc)₂R–vinblastine, contacts the side chain of Tyr β 224, which is sandwiched between the base of the nucleotide and ring D' (Fig. 1c). Taken together, these data strongly suggest that vinblastine binds to free tubulin in solution in a very similar manner to that observed in (Tc)₂R. The (Tc)₂R–vinblastine structure therefore provides a basis on which to rationalize both the binding of the drug to (Tc)₂R and its effects on tubulin.

We characterized the binding of vinblastine to (Tc)₂R in solution by monitoring fluorescence, taking advantage of tubulin–colchicine fluorescence changes upon vinblastine binding¹⁴. Our studies confirm the 1:1 vinblastine:(Tc)₂R stoichiometry observed in the structure (data not shown). From the linear dependence of the pseudo-first-order rate constants on vinblastine concentration (Fig. 3), values ($\pm$ s.d.) of $0.018 \pm 0.0016 \mu\text{M}^{-1} \text{s}^{-1}$ and $0.0063 \pm 0.0042 \text{s}^{-1}$ were derived for the association (k_+) and dissociation (k_-) rate constants, respectively. The value of k_+ is several orders of magnitude smaller

¹Laboratoire d'Enzymologie et Biochimie Structurales, UPR 9063, Centre National de la Recherche Scientifique, Bâtiment 34, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France. ²European Molecular Biology Laboratory (EMBL), Grenoble Outstation, 6 rue Jules Horowitz, BP 181, 38042 Grenoble Cedex 9, France. ³Institut de Chimie des Substances Naturelles, Centre National de la Recherche Scientifique, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France. ⁴Paul Scherrer Institut, Biomolecular Research, Structural Biology, CH-5232 Villigen PSI, Switzerland. ⁵U706 INSERM/UPMC, Institut du Fer à Moulin, 17 rue du Fer à Moulin, 75005 Paris, France. [†]Present address: Laboratoire Structure et Reconnaissance des Biomolécules, EA3637, Université Evry-Val d'Essonne, rue du Père Jarlan, 91025 Evry, France.

than expected for a diffusion-limited bimolecular reaction. This indicates that vinblastine binds to $(Tc)_2R$ in two consecutive steps, namely formation of a rapid equilibrium collision complex followed by a slower rearrangement linked to the fluorescence change that might coincide with the changes we have seen in the structure. The resulting equilibrium dissociation constant, K_d , is $0.4 \pm 0.3 \mu M$, close to the value derived for the binding of vinblastine to self-associated tubulin¹⁵, as might have been expected from the location of its binding site in $(Tc)_2R$.

A major effect of vinblastine is its capacity to induce the formation of spiral-like tubulin aggregates^{4,16}. As it interacts with both tubulins in $(Tc)_2R$, this immediately suggests a mechanism in which vinblastine crosslinks two tubulin molecules by interacting with the α -subunit of one and the β -subunit of the other. Such contacts would

axially stabilize a curved tubulin protofilament of the type observed with vinblastine-depolymerized microtubules¹⁷. Vinblastine also induces aggregation of the ternary complex of two tubulin heterodimers with the long C-terminal RB3-SLD α -helix ($(Tc)_2R_{hel}$) (Fig. 4a). According to our proposal, these aggregates are helical assemblies of $(Tc)_2R_{hel}$ complexes bridged by vinblastine through contacts with the $\alpha 1$ -tubulin and $\beta 2$ -tubulin subunits identical to those of the drug with $\alpha 2$ and $\beta 1$ in the $(Tc)_2R$ -vinblastine structure (Fig. 4b). Indeed, electron micrographs of $(Tc)_2R_{hel}$ -vinblastine reveal long curls of protofilaments (Fig. 4b), reminiscent of the more flexible and often shorter structures obtained with tubulin-colchicine under the same conditions (Supplementary Fig. 1). Remarkably, the radius of curls of $(Tc)_2R_{hel}$ on the grid, 23 ± 3 nm, is very similar to that (22 nm) of the helical atomic model we propose. This radius, which reflects the effects of RB3-SLD¹⁸ and of vinblastine, is close to the 21-nm radius of oligomers observed during microtubule depolymerization¹⁹ and is within the 12–25-nm range of the radii of tubulin curved assemblies induced by drugs or proteins (see refs 11, 20 and references therein).

In contrast with tubulin and $(Tc)_2R_{hel}$, vinblastine does not aggregate $(Tc)_2R$ (Fig. 4a). The only difference between $(Tc)_2R_{hel}$ and $(Tc)_2R$ is that the former complex lacks the N-terminal region of RB3-SLD. This region inhibits longitudinal protofilament-like tubulin contacts by capping the α -tubulin subunit, either when embedded in an SLD^{3,21} or autonomously (M. J. Clément *et al.*, unpublished observations). It also masks vinblastine-contacting tubulin residues that are required for drug-induced head-to-tail aggregation. In $(Tc)_2R$ these residues (Fig. 4c) are in the $\alpha 1$ subunit and their positions in the α -tubulin sequence overlap those of the $(Tc)_2R$ $\alpha 2$ subunit residues that contact vinblastine. They constitute a patch on the surface of α -tubulin (Fig. 4c) that is shared by the

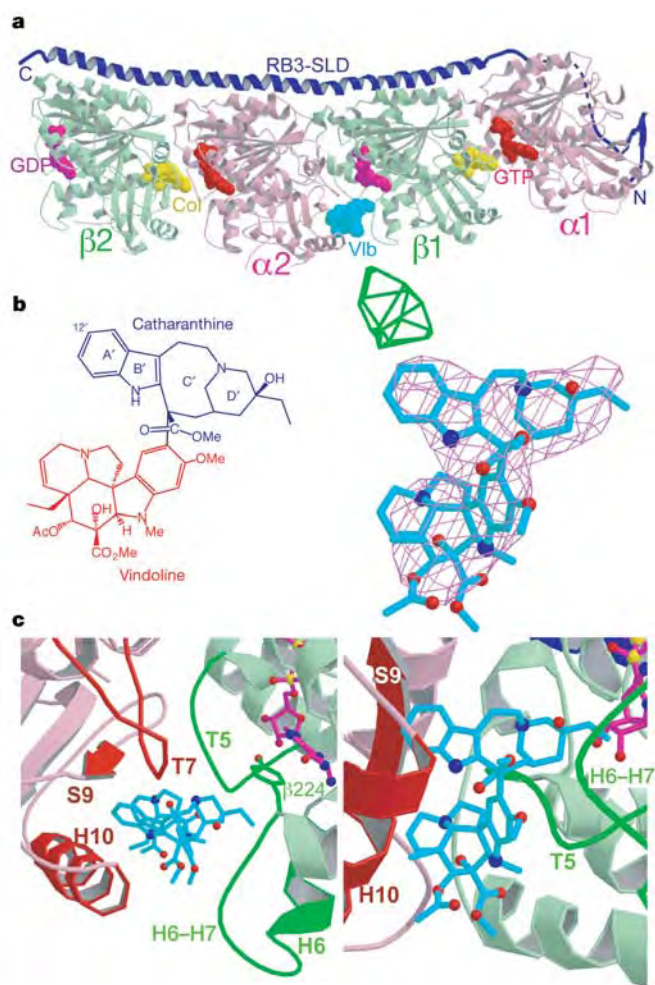


Figure 1 | The vinblastine-binding site. **a**, The location of vinblastine in $(Tc)_2R$. $(Tc)_2R$ -bound vinblastine (Vlb, cyan) is shown as a space-filling model; the complex consists of RB3-SLD and two tubulin $\alpha\beta$ heterodimers, with colchicine (Col, yellow) bound to the β -subunits at the interface with the α -subunit. **b**, Left: chemical formula of vinblastine. Right: σ_a -weighted $F_{obs} - F_{calc}$ omit map of vinblastine-soaked $(Tc)_2R$ crystals contoured at 3.5σ (magenta) overlapped with vinblastine; the difference electron density map between a C12'-bromovinblastine soaked $(Tc)_2R$ crystal after a low and a high dose of X-ray irradiation is shown in green (contoured at 5.5σ). **c**, Two perpendicular views of the vinblastine site. The labelled structural elements belong to the two domains that, together with a C-terminal helical hairpin, constitute the tubulin subunits. Loop T5 is part of the N-terminal $\beta 1$ tubulin nucleotide-binding domain. Helix H6 and loop H6-H7 link this domain to the intermediate domain. Helix H10, strand S9 and loop T7 are part of the $\alpha 2$ tubulin intermediate domain. Loops T5 and T7 contact the nucleotide in straight protofilaments.

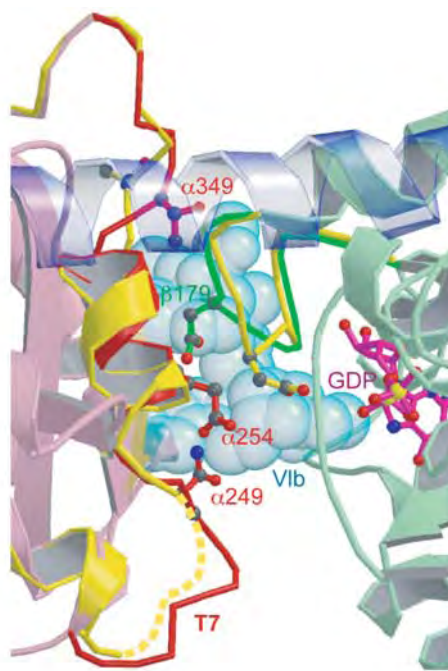


Figure 2 | Structural changes in $(Tc)_2R$ after binding vinblastine. The $\alpha 2$ and $\beta 1$ subunits of $(Tc)_2R$ are superimposed on the corresponding subunits of the $(Tc)_2R$ -vinblastine complex. The parts of $(Tc)_2R$ that do not coincide with $(Tc)_2R$ -vinblastine are in yellow. In $(Tc)_2R$, part of the T7 loop (dashed line) is disordered. Residue $\alpha 254$ has been proposed to enhance the tubulin GTPase activity⁹. Its distance to the $\beta 1$ subunit nucleotide is 11 Å; this indicates that the GTPase activity of the $\beta 1$ subunit might be inhibited in the $(Tc)_2R$ -vinblastine complex.

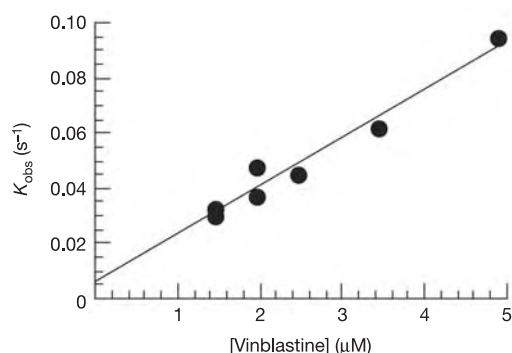


Figure 3 | Kinetics of vinblastine binding to (Tc)₂R. The pseudo-first-order plot of the apparent rate constants for binding of vinblastine at the indicated concentrations to 0.3 μM (Tc)₂R yields the values of k_+ (slope) and k_- (ordinate intercept).

binding sites for the RB3-SLD N-terminal cap and vinblastine. This patch is at the bottom of a shallow groove bordered by strand S9 and helix H10. It is hydrophobic, located at a longitudinal contact in microtubules⁸ and defines a distinct target on the tubulin surface for drugs aimed at interfering with microtubule assembly.

Vinblastine binds to microtubule tips²² and, at low concentrations, suppresses the dynamic instability of plus ends²³. This probably accounts for the mitotic block in cells exposed to clinically relevant doses of this drug²⁴. When used at much higher concentrations, vinblastine depolymerizes microtubules, giving rise in particular to protofilament spirals and curls (see ref. 4 and references therein). The (Tc)₂R–vinblastine structure suggests the following explanations of these effects. Microtubule stability requires both longitudinal

and lateral interactions between tubulin molecules, M loops of straight tubulin subunits being the central elements of the latter⁸. When a vinblastine molecule binds to microtubule tips, it modifies the longitudinal interface of the tubulin molecules that it contacts and constrains them to a curved assembly, because straight assemblies would clash sterically with vinblastine (Supplementary Fig. 2). In the simplest scheme, the vinblastine-contacting tubulin closer to the core of the microtubule is unaffected, whereas the tubulin closer to the end of the protofilament orients so that the longitudinal interface between these two molecules is similar to that in (Tc)₂R–vinblastine. The M loops of the latter heterodimer are substantially displaced in comparison with straight protofilaments. This displacement, estimated by comparing the structures of two straight tubulins assembled head-to-tail as in (Tc)₂R–vinblastine or in protofilaments², is larger than 10 Å (Supplementary Fig. 3). Therefore vinblastine, on binding, non-competitively interferes with tubulin's lateral interactions and competitively inhibits its longitudinal interactions that give rise to straight protofilaments, thus inhibiting its function. Because the binding of one drug molecule per microtubule end²² suppresses dynamic instability, it is tempting to conclude that, for this instability to occur, the integrity of microtubule inter-dimer contacts within and between protofilaments is required. At high concentrations of vinblastine, the proportion of missing microtubule-specific contacts increases further, which leads to microtubule disassembly.

Vinblastine and colchicine, the two prototypical examples of microtubule-depolymerizing drugs whose mechanisms have been structurally characterized (this study and ref. 3), introduce curvature into tubulin and tubulin assemblies. They do this at different sites, vinblastine at inter-dimer interfaces and colchicine at intra-dimer interfaces. Similar curved conformations have also been observed in the complex of tubulin with SLDs^{18,21} and with the long C-terminal

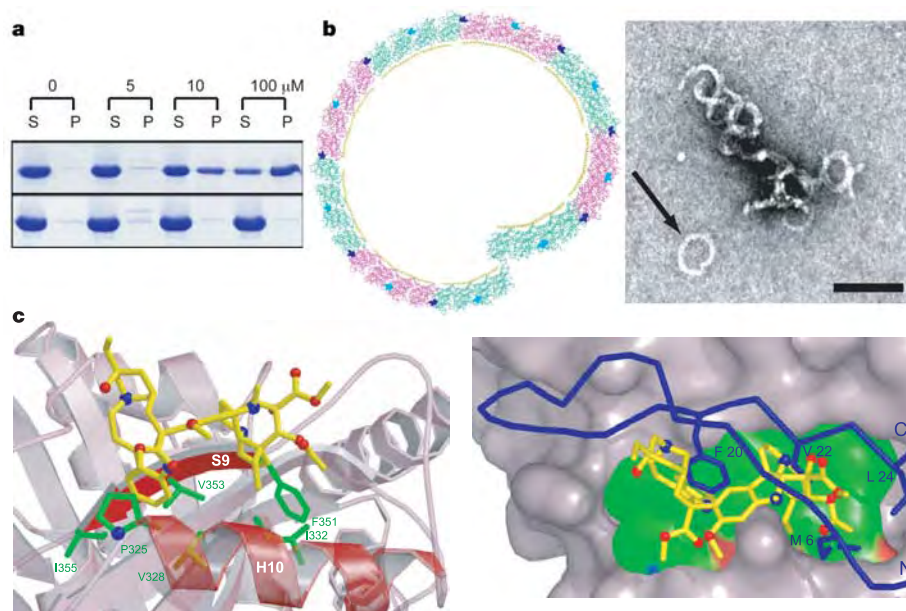


Figure 4 | The influence of SLDs on vinblastine-mediated formation of spiral aggregates. **a**, The variation in aggregation of (Tc)₂Rhel (top) and (Tc)₂R (bottom) as a function of vinblastine concentration was assayed by analysing the tubulin content in the supernatant (S) and pellet (P) after centrifugation of mixtures with vinblastine at the indicated concentrations, as described in Methods. Rhel (or RB3) is associated with tubulin (not shown). **b**, Left: atomic model of (Tc)₂Rhel aggregates based on the (Tc)₂R–vinblastine structure (see the text); the vinblastine molecules that crosslink two (Tc)₂Rhel complexes are in dark blue, Rhel is in yellow. Right: electron micrograph of negatively stained (Tc)₂Rhel–vinblastine aggregates;

the arrow indicates an isolated curl. Scale bar, 100 nm. **c**, Left: vinblastine (yellow) is positioned with respect to $\alpha 1$ identically to the way in which it contacts $\alpha 2$ in (Tc)₂R. The $\alpha 1$ residues shown (green) constitute the hydrophobic patch common to the binding sites of RB3-SLD N-terminal cap and vinblastine. Right: vinblastine (positioned as in the left panel) and the RB3-SLD N-terminal region (dark blue; residues 4–24, stathmin numbering) bound to (Tc)₂R are shown. In the surface view of (Tc)₂R, the hydrophobic patch is in green. Side chains of RB3-SLD residues that contact this patch are shown.

helix of stathmin²¹, which both sequester tubulin^{25,26}. Another class of proteins, internal motor domain kinesins, which depolymerize microtubules by using ATP as an energy source, have been modelled to bind preferentially to curved protofilaments²⁷. Forcing or fixing a curved tubulin assembly and/or conformation therefore represents an important component of the mechanisms used by both endogenous and exogenous tubulin ligands to disassemble microtubules or to prevent their assembly. It is also of interest that, in contrast with the binding sites for Taxol² and colchicine³, which are mostly embedded in a single tubulin heterodimer, the vinblastine binding site is shared equally between two heterodimers. Such interfacial interference has been described in rare cases (see ref. 28 and references therein), but it might be particularly suited to tubulin because of the latter's propensity to self-associate.

METHODS

Proteins and buffers. These are described in Supplementary Information.

Structure determination. Crystals of the (Tc)₂R complex³ were soaked for 12–24 h with a 2 mM solution of vinblastine before being flash-frozen in liquid nitrogen. These crystals diffracted up to 4 Å resolution. The (Tc)₂R structure³ was used as a starting model and was refined further as in ref. 3. The strongest feature of the extra density in an $F_{\text{obs}} - F_{\text{calc}}$ omit map was attributed to one vinblastine molecule. Vinblastine was refined as a rigid body. Initial topology parameters for the vinblastine structure²⁹ were generated using the program PRODRG³⁰. Statistics for data processing and refinement are summarized in Supplementary Table 1.

To ascertain the vinblastine orientation, crystals were soaked with 12'-bromovinblastine⁵. We collected two low-resolution (6 Å) data sets with an attenuated beam; between obtaining the two data sets we 'burned' the crystal for 30 s with X-rays at maximum flux. The first peak in a difference Fourier map between data collected before and after the burn corresponds to the loss of the bromide atom. Most of the following peaks are also consistent with damage induced by X-rays and correspond in particular to the decarboxylation of Asp and Glu residues.

Vinblastine binding. Quenching of (Tc)₂R fluorescence after binding of vinblastine amounted to up to about 10% (λ_{ex} 280 nm; λ_{em} 340 nm). To a 0.3 μ M (Tc)₂R solution we added at least a fivefold excess of vinblastine, which made the binding reaction pseudo-first-order. Because vinblastine does not induce (Tc)₂R self-association, the binding data were analysed as a single ligand-receptor process. The exponential decrease of the fluorescence signal was fitted with the equation

$$[C] = [T](1 - \exp(-k_{\text{obs}}t))$$

where [C] and [T] are the concentrations of (Tc)₂R–vinblastine and total (Tc)₂R, respectively. The plot of apparent rate constant k_{obs} at different vinblastine concentrations yields both the association rate k_+ (slope) and the dissociation rate k_- (ordinate intercept), from which K_d is derived.

Vinblastine-mediated aggregation. (Tc)₂Rhel and (Tc)₂R complexes (2.5 μ M) were incubated with various vinblastine concentrations at room temperature (20–25 °C) for 30 min. After high-speed centrifugation (250,000g for 15 min), the supernatant and pellet were analysed by SDS–polyacrylamide gel electrophoresis.

Electron microscopy of vinblastine-induced tubulin and (Tc)₂Rhel assemblies. (Tc)₂Rhel complex (10 μ M) or tubulin–colchicine (10 μ M) was incubated with 50 μ M vinblastine for 30 min at room temperature. Sample aliquots of 5 μ l were applied to a weakly glow-discharged carbon-coated 400-mesh/inch copper grid, left to adsorb for 30 s, washed twice with water, and negatively stained for 20 s with 2% (w/v) uranyl acetate. Specimens were examined in a Philips Morgagni transmission electron microscope operated at 80 kV. Micrographs were recorded with a Megaview III charge-coupled device camera at nominal magnifications of $\times 30,000$ and $\times 50,000$.

Received 6 January; accepted 21 March 2005.

- Jordan, M. A. & Wilson, L. Microtubules as a target for anticancer drugs. *Nature Rev. Cancer* **4**, 253–265 (2004).
- Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature* **391**, 199–203 (1998).
- Ravelli, R. B. G. *et al.* Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain. *Nature* **428**, 198–202 (2004).
- Himes, R. H. Interactions of the catharanthus (Vinca) alkaloids with tubulin and microtubules. *Pharmacol. Ther.* **51**, 257–267 (1991).
- Trouet, A. B., Hannart, J. A. A. & Rao, K. S. B. (Omnichem S.A., Belg), US Patent 4,639,456 (1987).

- Ennifar, E., Carpentier, P., Ferrer, J. L., Walter, P. & Dumas, P. X-ray-induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing. *Acta Crystallogr. D* **58**, 1262–1268 (2002).
- Ravelli, R. B. G., Leiros, H. K., Pan, B., Caffrey, M. & McSweeney, S. Specific radiation damage can be used to solve macromolecular crystal structures. *Structure* **11**, 217–224 (2003).
- Nogales, E., Whittaker, M., Milligan, R. A. & Downing, K. H. High-resolution model of the microtubule. *Cell* **96**, 79–88 (1999).
- Nogales, E., Downing, K. H., Amos, L. A. & Löwe, J. Tubulin and FtsZ form a distinct family of GTPases. *Nature Struct. Biol.* **5**, 451–458 (1998).
- Rai, S. S. & Wolff, J. Localization of the vinblastine-binding site on β -tubulin. *J. Biol. Chem.* **271**, 14707–14711 (1996).
- Mitra, A. & Sept, D. Localization of the antimetabolic peptide and depsipeptide binding site on β -tubulin. *Biochemistry* **43**, 13955–13962 (2004).
- Usui, T. *et al.* The anticancer natural product pironetin selectively targets Lys352 of α -tubulin. *Chem. Biol.* **11**, 799–806 (2004).
- Lober, S. *et al.* Vinca alkaloid-induced tubulin spiral formation correlates with cytotoxicity in the leukemic L1210 cell line. *Biochemistry* **39**, 12053–12062 (2000).
- Sackett, D. L. Vinca site agents induce structural changes in tubulin different from and antagonistic to changes induced by colchicine site agents. *Biochemistry* **34**, 7010–7019 (1995).
- Na, G. C. & Timasheff, S. N. Interaction of vinblastine with calf brain tubulin. *Biochemistry* **25**, 6214–6222 (1986).
- Weisenberg, R. C. & Timasheff, S. N. Aggregation of microtubule subunit protein. Effects of divalent cations, colchicine and vinblastine. *Biochemistry* **9**, 4110–4116 (1970).
- Haskins, K. M., Donoso, J. A. & Himes, R. H. Spirals and paracrystals induced by Vinca alkaloids: evidence that microtubule-associated proteins act as polycations. *J. Cell Sci.* **47**, 237–247 (1981).
- Gigant, B. *et al.* The 4 Å X-ray structure of a tubulin:stathmin-like domain complex. *Cell* **102**, 809–816 (2000).
- Müller-Reichert, T., Chrétien, D., Severin, F. & Hyman, A. A. Structural changes at microtubule ends accompanying GTP hydrolysis: Information from a slowly hydrolyzable analogue of GTP, guanylyl (α,β)methylene diphosphonate. *Proc. Natl Acad. Sci. USA* **95**, 3661–3666 (1998).
- Nogales, E., Wang, H. W. & Niederstrasser, H. Tubulin rings: which way do they curve? *Curr. Opin. Struct. Biol.* **13**, 256–261 (2003).
- Steinmetz, M. O. *et al.* Op18/stathmin caps a kinked protofilament-like tubulin tetramer. *EMBO J.* **19**, 572–580 (2000).
- Wilson, L., Jordan, M. A., Morse, A. & Margolis, R. L. Interaction of vinblastine with steady-state microtubules *in vitro*. *J. Mol. Biol.* **159**, 125–149 (1982).
- Toso, R. J., Jordan, M. A., Farrell, K. W., Matsumoto, B. & Wilson, L. Kinetic stabilization of microtubule dynamic instability *in vitro* by vinblastine. *Biochemistry* **32**, 1285–1293 (1993).
- Jordan, M. A., Thrower, D. & Wilson, L. Mechanism of inhibition of cell proliferation by Vinca alkaloids. *Cancer Res.* **51**, 2212–2222 (1991).
- Charbaut, E. *et al.* Stathmin family proteins display specific molecular and tubulin binding properties. *J. Biol. Chem.* **276**, 16146–16154 (2001).
- Redeker, V. *et al.* Probing the native structure of stathmin and its interaction domains with tubulin. *J. Biol. Chem.* **275**, 6841–6849 (2000).
- Ogawa, T., Nitta, R., Okada, Y. & Hirokawa, N. A common mechanism for microtubule stabilizers-M type kinesins stabilize curling of the protofilament using the class-specific neck and loops. *Cell* **116**, 591–602 (2004).
- Renault, L., Guibert, B. & Cherfils, J. Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. *Nature* **426**, 525–530 (2003).
- Bau, R. & Jin, K. K. Crystal structure of vinblastine. *J. Chem. Soc. Perkin Trans. 1*, 2079–2082 (2000).
- Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements F.R. thanks F. Guéritte and D. Guénard for support. We thank M.-F. Carlier for help and discussions. We are grateful to the Interdisciplinary Center for Microscopy of the University of Basel for access to the transmission electron microscope. This work was supported by the Association pour la Recherche sur le Cancer, the Centre National pour la Recherche Scientifique and the Institut National pour la Santé et la Recherche Médicale. Diffraction data were collected on beamline ID14-4 at European Synchrotron Radiation Facility. C.W. was supported in part by the K.C. Wong foundation and by an ARC postdoctoral fellowship.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession number 1Z2B (tubulin–colchicine:RB3–SLD–vinblastine). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.K. (knossow@lebs.cnrs-gif.fr).

Insights into microtubule nucleation from the crystal structure of human γ -tubulin

Hector Aldaz^{1*†}, Luke M. Rice^{1*}, Tim Stearns² & David A. Agard¹

Microtubules are hollow polymers of $\alpha\beta$ -tubulin that show GTP-dependent assembly dynamics and comprise a critical part of the eukaryotic cytoskeleton. Initiation of new microtubules *in vivo* requires γ -tubulin, organized as an oligomer within the 2.2-MDa γ -tubulin ring complex (γ -TuRC) of higher eukaryotes^{1–3}. Structural insight is lacking regarding γ -tubulin, its oligomerization and how it promotes microtubule assembly. Here we report the 2.7-Å crystal structure of human γ -tubulin bound to GTP- γ S (a non-hydrolysable GTP analogue). We observe a ‘curved’ conformation for γ -tubulin–GTP- γ S, similar to that seen for GDP-bound, unpolymerized $\alpha\beta$ -tubulin⁴. Tubulins are thought to represent a distinct class of GTP-binding proteins, and conformational switching in γ -tubulin might differ from the nucleotide-dependent switching of signalling GTPases. A crystal packing interaction replicates the lateral contacts between α - and β -tubulins in the microtubule⁵, and this association probably forms the basis for γ -tubulin oligomerization within the γ -TuRC. Laterally associated γ -tubulins in the γ -TuRC might promote microtubule nucleation by providing a template that enhances the intrinsically weak lateral interaction between $\alpha\beta$ -tubulin heterodimers. Because they are dimeric, $\alpha\beta$ -tubulins cannot form microtubule-like lateral associations in the curved conformation⁵. The lateral array of γ -tubulins we observe in the crystal reveals a unique functional property of a monomeric tubulin.

The X-ray crystal structure of a human γ -tubulin–GTP- γ S complex was determined by molecular replacement at 2.7-Å resolution (see Methods) (Fig. 1a, b and Supplementary Table 1). This is the first structure of a monomeric tubulin and the highest resolution structure of any tubulin to date. We also determined the essentially identical structure of a γ -tubulin–GTP complex at somewhat lower resolution (see Supplementary Methods), providing strong evidence that the γ -tubulin–GTP- γ S structure represents a bona fide GTP-bound conformation of γ -tubulin. These are the first tubulin structures with GTP or a GTP analogue bound at an exchangeable site.

As expected from the high sequence conservation within the tubulin superfamily, the overall structure of γ -tubulin is similar to previously reported α - and β -tubulin structures^{4,6}. Relative to α - and β -tubulin, there are several sites at which γ -tubulin has insertions or deletions that are highly conserved across species (see Supplementary Fig. 1). Two of these, the insertion two Gly residues at position 99 in the T3 loop and the Asp insertion at position 175, are located in regions of disorder in the γ -tubulin structure, but are located on a presumptive longitudinal contact surface and might therefore help to determine γ -tubulin-specific longitudinal interaction properties (nomenclature and interaction surfaces are defined as in refs. 7, 8. Briefly, longitudinal interactions occur along a protofilament; lateral

interactions occur between protofilaments, around the microtubule wall). His 107 is also lacking in γ -tubulin; this deletion restores a more ideal α -helical register to helix H3' and removes a bulge from this region of the structure (see Supplementary Fig. 1c). Our structure reveals that removal of this bulge changes a lateral interaction surface of γ -tubulin, causing the amino-terminal end of helix H3 to protrude (see Supplementary Fig. 1c). Several other features unique to γ -tubulin sequences reside on surfaces equivalent to those that make longitudinal or lateral contacts between α - and β -tubulin in microtubules (see Supplementary Fig. 1a, b). It is difficult to determine the relative importance of some of these γ -tubulin-specific features, because the molecular interactions underlying lateral interactions in microtubules have yet to be visualized in atomic detail. Nevertheless, it is striking that these and other unique features of the γ -tubulin family map to known tubulin–tubulin interaction surfaces. These residues might collectively dictate γ -tubulin interaction strengths distinct from those of α - or β -tubulin.

In contrast to these local surface differences, many of the elements responsible for GTP binding are shared between α -, β -, and γ -tubulin^{4,6,7}. Similar to β -tubulin, the guanine base of GTP in γ -tubulin is sandwiched between Phe 225 and Cys 13, with Asn 229 and Asn 207 contributing important hydrogen bonds (Fig. 1b). With the exception of conserved Ser 140 and Thr 145 residues, most of the phosphate contacts are made by backbone amides: Cys 13 to O1 α , Gln 12, Gly 144, Gly 146 and Thr 145 to O1 β , and Gly 144 and Thr 145 to O3 γ . We have placed an Mg²⁺ ion in electron density near the β - and γ -phosphates, making contacts with phosphate oxygens O2 β and O3 β , and with Asp 68, Glu 70 and a water molecule.

Not all features of the nucleotide binding pockets are identical between γ - and β -tubulin, however. In addition to the disordered T3 loop mentioned above, the T5 loop is also disordered in γ -tubulin. In $\alpha\beta$ -tubulin, the T3 and T5 loops participate in longitudinal contacts, so it is possible that they only become ordered and form interactions with GTP/GDP upon longitudinal tubulin association. We performed comparative nucleotide binding studies to determine the degree to which structural similarity between the nucleotide-binding pockets of γ - and β -tubulin extends to shared GTP-binding properties (Fig. 1c). γ - and β -tubulin both bind GTP with very similar affinity (~60 nM). Competition experiments using GDP show that both γ - and β -tubulin have a similar preference for GTP over GDP. Together with strong conservation of sequence and structure, the essentially identical nucleotide-binding affinities of γ -tubulin and β -tubulin indicate that they probably share common nucleotide-binding energetics that are largely unaffected by structural differences in the T3 and T5 loops. This also makes it likely that, analogous to its action on β -tubulin, α -tubulin can stimulate the GTPase activity of γ -tubulin upon longitudinal association.

¹Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, 600 16th Street, San Francisco, California 94143, USA.

²Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA. [†]Present address: Department of Molecular and Cell Biology, University of California, Berkeley, 16 Barker Hall, Berkeley, California 94720-3202, USA.

*These authors contributed equally to this work.

Two distinct conformations of $\alpha\beta$ -tubulin have been characterized structurally. The differences between the two conformations involve small but significant domain rearrangements within and between both tubulin subunits. The 'straight' $\alpha\beta$ -tubulin conformation has only been observed in microtubules and in the

protofilament-containing sheets induced by polymerization in the presence of zinc^{5,6}. A 'curved' conformation has been observed in unpolymerized $\alpha\beta$ -tubulin bound to the stathmin-like domain of the microtubule-destabilizing protein RB3. One feature that distinguishes these two tubulin conformations is the arrangement of the segment containing helices H6 and H7 (ref. 4). Another is the relative orientation between the structurally conserved, rigid N-terminal domain and the more variable intermediate domain⁴. Notably, superposition of the straight and curved β -tubulin conformations onto the γ -tubulin structure, using the structurally well-conserved N-terminal domain, reveals that the arrangement of indicative domains in the γ -tubulin-GTP γ S structure more closely resembles that of the curved conformation (Fig. 2, Table 1 and Supplementary Fig. 2). Although the conformational match is extensive, it is not perfect. The surface-exposed helices H9 and H10 of γ -tubulin are distinctively arranged, differing from the arrangement seen in the two previously determined $\alpha\beta$ -tubulin structures (Fig. 2, Table 1 and Supplementary Fig. 2). However, most secondary structural elements in γ -tubulin are much closer to their curved $\alpha\beta$ -tubulin counterparts. Using this superposition, atoms from helices H6 through H7 in γ -tubulin have an r.m.s. coordinate deviation of 2.46 Å to the straight conformation of β -tubulin, but only 1.38 Å to the curved conformation of β -tubulin. Similarly, atoms in the largely buried β -sheet of the intermediate domain of γ -tubulin have an r.m.s. coordinate deviation of 2.31 Å to the straight conformation and only 1.18 Å to the curved conformation. Comparisons to the two known α -tubulin conformations yield similar results (Table 1 and Supplementary Fig. 2).

When bound to either GTP- γ S or GTP, γ -tubulin adopts a curved conformation with a characteristic arrangement of helices H6 and H7 and the intermediate domain. The similarity between this conformation and that seen for both curved α - and β -tubulin in the unpolymerized, RB3-bound state (Fig. 2, Table 1 and Supplementary Fig. 2) suggests that this is a functionally relevant tubulin conformation. The arrangement of helices H6 and H7 in the curved conformation of α - and β -tubulin is incompatible with straight longitudinal interactions⁴ like those presumed to occur between the γ -TuRC and a microtubule (see below). Thus, our unexpected observation of a curved, GTP-bound conformation represents the first structural evidence that curved-to-straight transitions might occur in γ -tubulin at the minus-end of a microtubule, and that they might be substantially nucleotide-independent. This would distinguish conformational switching in γ -tubulin from that in signalling GTPases, consistent with the observation that tubulins represent a distinct class of GTP-binding proteins⁸. In addition, several recent structures of the monomeric prokaryotic tubulin homologue FtsZ in different nucleotide states all show essentially the same curved conformation⁹.

One of the observed crystal packing interactions bears a striking similarity to the homotypic lateral interactions previously only seen to occur in the microtubule lattice⁵ (Fig. 3). The γ -tubulin interaction lacks the lateral curvature of the cylindrical microtubule lattice, but uses a virtually identical contact region and buries a comparable surface area (Fig. 3). The γ -tubulin structure thus shows that, at the level of a tubulin monomer, a curved conformation can form microtubule-like lateral interactions. In contrast, the curved conformation of the $\alpha\beta$ -tubulin heterodimer discourages microtubule growth for two reasons: it disrupts the parallel presentation of α - and β -tubulin lateral interaction surfaces within a heterodimer, and the positions of helix H6 and the H6-H7 loop are incompatible with straight longitudinal assembly between heterodimers⁴. By virtue of being monomeric, γ -tubulin acquires a unique functional property—the ability to self-associate laterally while in the curved conformation.

Lateral interactions between $\alpha\beta$ -tubulins must be sufficiently flexible to permit microtubules formed *in vitro* to contain between 9 and 16 protofilaments, and to allow the formation of flat sheets that

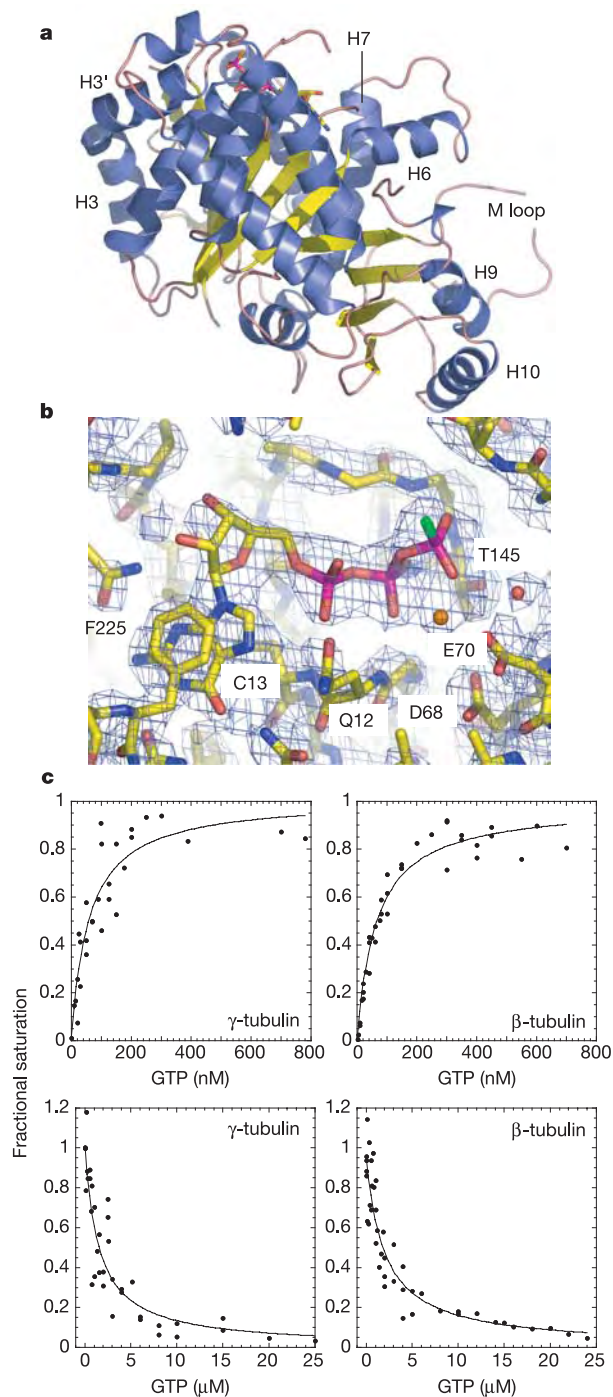


Figure 1 | Structure and nucleotide binding-properties of γ -tubulin. **a**, Cartoon representation of the γ -tubulin structure. **b**, Representative electron density from a complete annealed σ_A -weighted $2mF_o - DF_c$ map computed from the final model. GTP- γ S and interacting residues are shown. **c**, Nucleotide binding studies reveal that γ -tubulin (left) and the exchangeable site on β -tubulin (right) have similar GTP/GDP binding properties. The affinities of γ -tubulin and β -tubulin for GTP are 58.4 ± 12.6 nM and 64.5 ± 6.3 nM (respectively) and GDP affinities are 1.13 ± 0.2 μ M for γ -tubulin and 1.55 ± 0.25 μ M for β -tubulin.

have been observed as intermediates in microtubule elongation¹⁰. The intrinsic flexibility of the structurally conserved lateral interaction, together with local sequence conservation (see Supplementary Fig. 1b, c), makes it likely that lateral interactions like those observed in the crystal form the basis for cylindrically curved arrays of γ -tubulin such as those found in the γ -TuRC. It also suggests that components of the γ -TuRC other than γ -tubulin have important roles in establishing a specific ring-like geometry.

Our results provide insight into the possible mechanisms and

regulation of γ -tubulin-mediated microtubule nucleation. The low stability of lateral associations between $\alpha\beta$ -tubulin heterodimers is believed to be the weak link in microtubule assembly. Appropriately spaced lateral assemblies of γ -tubulin, analogous to those we observe in the crystal (Fig. 3), and such as we propose exist within the γ -TuRC, have the potential to make strong longitudinal contacts with multiple $\alpha\beta$ -tubulins. These assemblies profit from the unique propensity of monomeric γ -tubulin to form stable lateral interactions, with spacing and pitch coincident with those of the

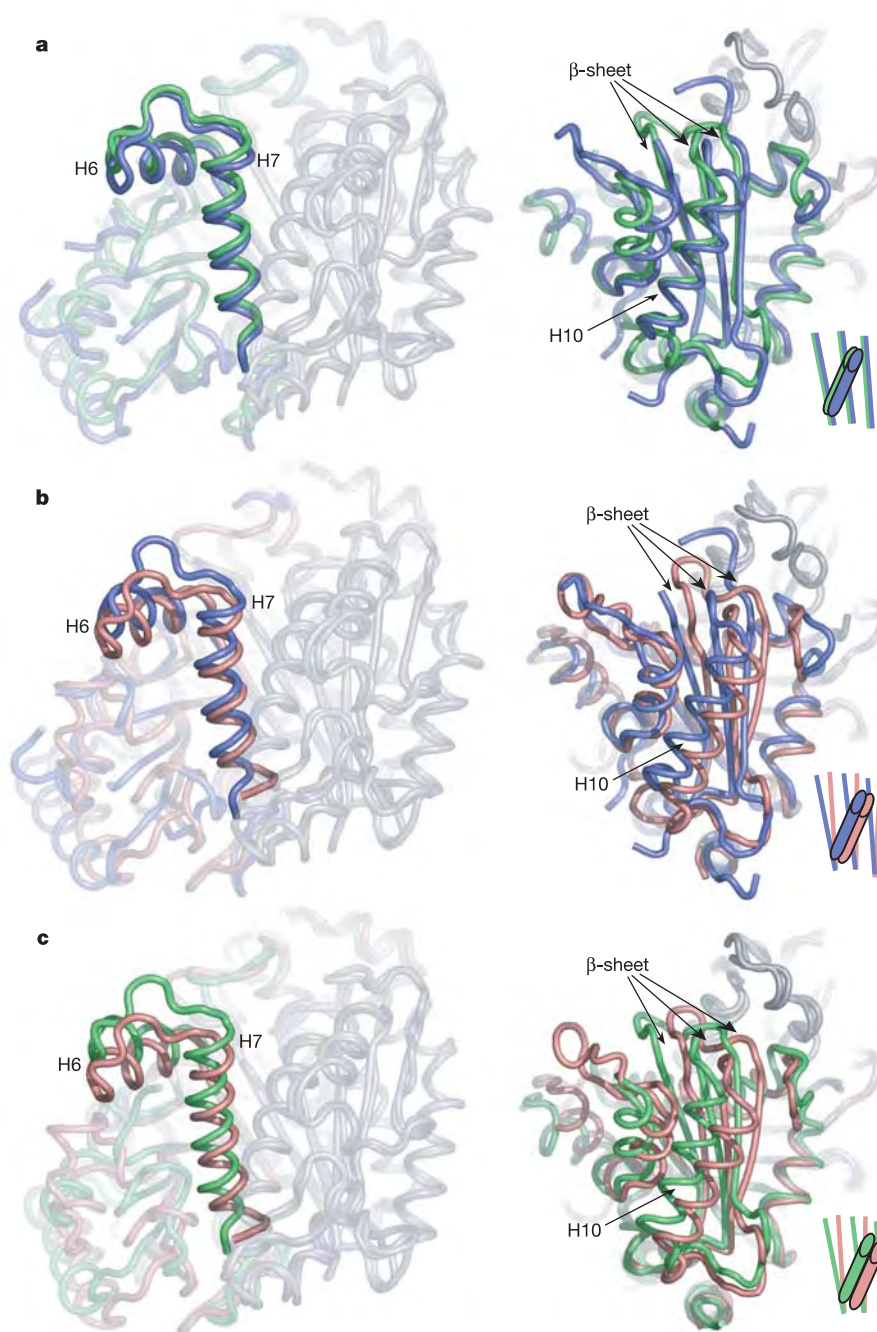


Figure 2 | γ -tubulin adopts a curved conformation. Structural superposition of β -tubulin conformations onto γ -tubulin using the rigid N-terminal domain. **a**, The γ -tubulin structure (blue) and the curved β -tubulin structure (green) share a similar arrangement of the H6–H7 segment (left) and of intermediate domains (right). **b**, The γ -tubulin structure (blue) and the straight β -tubulin structure (pink) show

characteristic differences in the orientation of the H6–H7 segment (left) and the intermediate domain (right). **c**, Comparison between the curved (green) and straight (pink) β -tubulin conformations, illustrating the characteristic differences in the H6–H7 segment (left) and the intermediate domain (right).

Table 1 | Statistics for superposition of α - and β -tubulin conformations onto γ -tubulin

Tubulin (α or β); straight (S) or curved (C)	Intermediate β -sheet				H10				H6-H7 segment				N-terminal domain			
	α (S)	α (C)	β (S)	β (C)	α (S)	α (C)	β (S)	β (C)	α (S)	α (C)	β (S)	β (C)	α (S)	α (C)	β (S)	β (C)
r.m.s. deviations ($\text{\AA}$)																
γ -tubulin	2.18	1.14	2.31	1.18	3.68	2.65	4.23	3.20	1.98	0.99	2.46	1.38	1.09	1.21	1.36	1.15
α -tubulin (S)	-	2.11	1.04	2.43	-	4.51	1.83	3.95	-	1.82	1.06	2.77	-	1.01	0.92	1.01
α -tubulin (C)	-	-	2.27	1.26	-	-	5.16	2.75	-	-	2.29	1.78	-	-	1.35	1.12
β -tubulin (S)	-	-	-	2.68	-	-	-	4.28	-	-	-	3.24	-	-	-	1.10

All known α - and β -tubulin conformations (straight, PDB code 1JFF; curved, PDB code 1SAO) were aligned to the γ -tubulin structure using the N-terminal domain (residues 1-175). Pairwise r.m.s. coordinate deviations for indicated secondary structural elements of the intermediate domain were calculated for these aligned models. The N-terminal domain r.m.s. deviations (rightmost four columns) are included as a control to show that this domain aligns well for all sequences and conformations. The intermediate β -sheet includes residues 267-273, 313-320, 352-358 and 373-38. Helix H10 includes residues 322-340. The H6-H7 segment includes residues 205-240. Deviations between curved and straight conformations of different sequences are significantly greater than deviations between like conformations and like sequences.

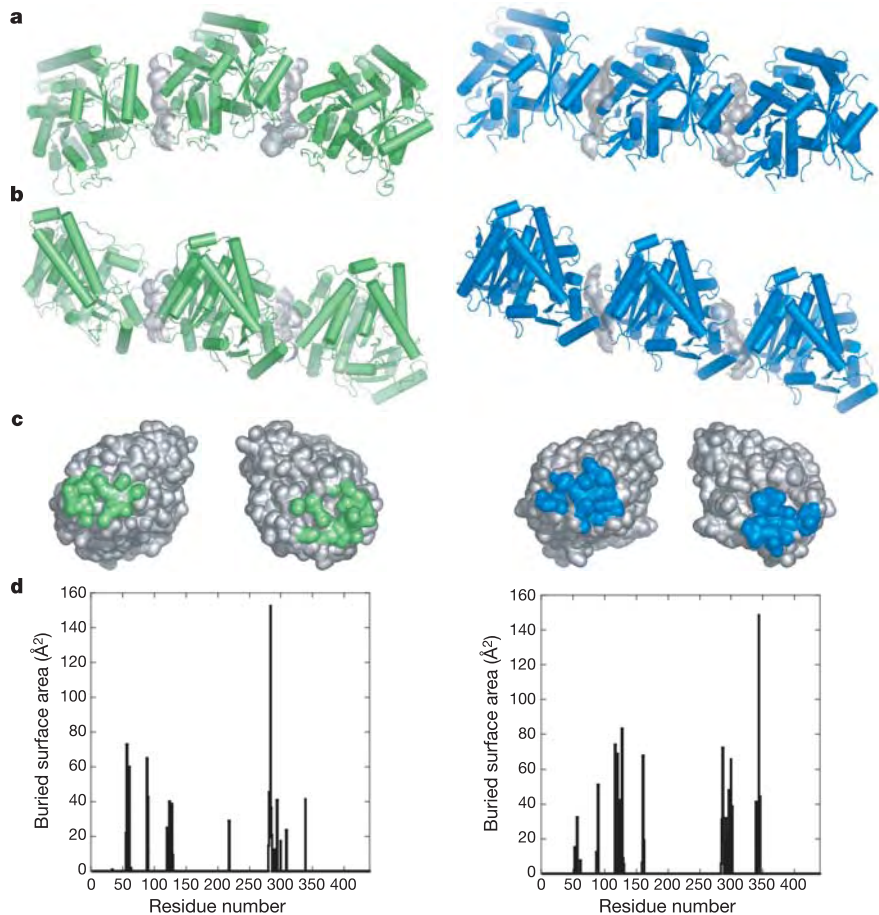


Figure 3 | Lateral interactions between γ -tubulins resemble lateral interactions in the microtubule lattice. **a**, ‘Minus end’ views of laterally interacting β -tubulins in the microtubule lattice (green) (K. Downing, personal communication), and laterally interacting γ -tubulins in the crystal (blue). Contact regions between monomers are indicated by the grey surfaces on the central monomer. **b**, Comparative ‘outside’ views of the same interactions, showing a similar pitch for both. **c**, Lateral interaction regions

of β -tubulin in the microtubule lattice (green) and γ -tubulin in the crystal (blue) are indicated on the molecular surface. Microtubule and γ -tubulin crystal interaction footprints are very similar. **d**, Comparison of the buried surface area (in $\text{\AA}^2$) at each position for β -tubulin lateral interactions in the microtubule lattice (left) and γ -tubulin crystal interactions (right). Virtually identical regions of the structure are involved in both interactions.

microtubule lattice. Presumably, these lateral assemblies of γ -tubulin promote microtubule assembly by providing a template that directly enhances lateral interactions between adjacent γ -TuRC-bound $\alpha\beta$ -tubulins. Whether γ -tubulin within the γ -TuRC is present in a curved or straight conformation remains to be determined. However, the significant increase in intermolecular contact surface area observed when $\alpha\beta$ -tubulin straightens in the microtubule lattice argues that γ -tubulin is also likely to switch conformations, either in response to α -tubulin binding or through the action of other γ -TuRC components. By analogy to β -tubulin, GTP binding and hydrolysis

on γ -tubulin might regulate its affinity for α -tubulin, potentially providing a mechanism for the observed release of microtubules from the centrosome¹¹ or tuning of the nucleating activity of γ -TuRC.

METHODS

Protein expression, crystallization and GTP/GDP binding experiments. See the Supplementary Information for details of expression, purification, GTP/GDP binding experiments, and protein crystallization. **Structure determination.** Diffraction data were collected at beamline 8.2.1

Table 2 | X-ray statistics

Refinement	GTP- γ S	GTP
Resolution (Å)	2.7	3.0
$R_{\text{work}}/R_{\text{free}}$	24.0/29.2	24.7/30.9
Number of atoms		
Protein	3145	3078
Ligand/ion	33	33
Water	22	0
B-factors		
Protein	52.8	49.4
Ligand/ion	44.8	46.4
Water	41.1	N/a
r.m.s deviations		
Bond lengths (Å)	0.0085	0.0086
Bond angles (°)	1.67	1.95

(HHMI) at the Advanced Light Source in Berkeley, California. Thin, radiation sensitive crystals required the merging of data from two crystals in order to obtain reasonably complete data sets at resolutions higher than 3.1-Å. Data processing and reduction were carried out using the HKL2000 package¹². Molecular replacement searches and subsequent refinements were carried out using CNS¹³, and model building was performed using O¹⁴. Initial phases for the GTP- γ S data were determined by molecular replacement using a β -tubulin search model⁴ (PDB accession number 1SA0) with side chains truncated to alanine and cofactors (GDP, Mg²⁺, colchicine and waters) removed. Regions of the search model not confirmed by the initial electron density map were deleted, and side chains with clear electron density in the initial map were added. As the model phases improved, additional side chains and loops were added when there was interpretable electron density. GTP- γ S was added only after the electron density clearly indicated the presence of the γ -phosphate. Over the course of model refinement and rebuilding, complete annealed omit maps were used extensively to help avoid model bias. The lower resolution GTP-bound structure was solved using the GTP- γ S structure stripped of cofactors as a search model, and subsequently refined in a similar manner to the GTP- γ S structure. Refinement statistics are presented in Table 2. Figures were made using PyMOL¹⁵.

Received 10 December 2004; accepted 31 March 2005.

1. Stearns, T. & Kirschner, M. *In vitro* reconstitution of centrosome assembly and function: the central role of γ -tubulin. *Cell* **76**, 623–637 (1994).
2. Zheng, Y. *et al.* Nucleation of microtubule assembly by a γ -tubulin-containing ring complex. *Nature* **378**, 578–583 (1995).
3. Moritz, M. *et al.* Recruitment of the γ -tubulin ring complex to *Drosophila* salt-stripped centrosome scaffolds. *J. Cell Biol.* **142**, 775–786 (1998).
4. Ravelli, R. B. G. *et al.* Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain. *Nature* **378**, 198–202 (2004).
5. Li, H. *et al.* Microtubule structure at 8 Å resolution. *Structure (Camb.)* **10**, 1317–1328 (2002).
6. Lowe, J. *et al.* Refined structure of $\alpha\beta$ -tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313**, 1045–1057 (2001).
7. Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature* **391**, 199–203 (1998).
8. Nogales, E. *et al.* Tubulin and FtsZ form a distinct class of GTPases. *Nature Struct. Biol.* **5**, 451–458 (1998).
9. Oliva, M. A., Cordell, S. C. & Lowe, J. Structural insights into FtsZ protofilament formation. *Nature Struct. Mol. Biol.* **11**, 1243–1250 (2004).
10. Chretien, D., Fuller, S. D. & Karsenti, E. Structure of growing microtubule ends: two-dimensional sheets close into tubes at variable rates. *J. Cell Biol.* **129**, 1311–1328 (1995).
11. Keating, T. J. *et al.* Microtubule release from the centrosome. *Proc. Natl Acad. Sci. USA* **94**, 5078–5083 (1997).
12. Otwinowski, Z. & Minor, W. in *Methods in Enzymology: Macromolecular Crystallography, Part A* (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic, New York, 1997).
13. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
14. Jones, T. A. *et al.* Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
15. Delano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, California, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Murphy for protein purification advice early in the project, the Mullins laboratory for the use of their ultraviolet illuminator, and the Agard laboratory centrosome group for discussions and input. We acknowledge the support of this work by grants from the National Institutes of Health (D.A.A. and T.S.) and the Howard Hughes Medical Institute. H.A. acknowledges support from an NIGMS predoctoral fellowship, and L.M.R. was a Paul Sigler/Agouon Institute fellow of the Helen Hay Whitney Foundation. D.A.A. is a Howard Hughes Medical Institute Investigator.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 1Z5V and 1Z5W. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.A. (agard@msg.ucsf.edu).

ERRATUM

doi:10.1038/nature03683

Reduction of hysteresis losses in the magnetic refrigerant $Gd_5Ge_2Si_2$ by the addition of iron

Virgil Provenzano, Alexander J. Shapiro & Robert D. Shull

Nature 429, 853–857 (2004)

The date of acceptance of this Letter was 17 May 2004, and not 5 December 2004 as published.

ERRATUM

doi:10.1038/nature03684

Ecological constraints on diversification in a model adaptive radiation

Rees Kassen, Martin Llewellyn & Paul B. Rainey

Nature 431, 984–988 (2004)

This Letter was submitted on 17 September 2003 and accepted on 9 August 2004; the dates are incorrect as published.

ERRATUM

doi:10.1038/nature03621

Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults

Jeffrey J. McGuire, Margaret S. Boettcher & Thomas H. Jordan

Nature 434, 457–461 (2005)

This Article was accepted for publication on 19 January 2005, and not on 19 December 2005 as published.

CORRIGENDUM

doi:10.1038/nature03656

A universal trend of amino acid gain and loss in protein evolution

I. King Jordan, Fyodor A. Kondrashov, Ivan A. Adzhubei, Yuri I. Wolf, Eugene V. Koonin, Alexey S. Kondrashov & Shamil Sunyaev

Nature 433, 633–638 (2005)

We reported a universal trend of amino-acid gain and loss observed for recent evolutionary history among a diverse set of 15 taxa, with amino acids of declining frequencies being the first to be incorporated into the genetic code and those of increasing frequencies being late recruits. We have since discovered that a similar scenario for protein evolution was proposed by Zuckerkandl and colleagues more than thirty years ago¹. Their analysis of a far smaller vertebrate-specific data set of two protein families also revealed asymmetric patterns of amino-acid substitution, and they went on to speculate that “extrapolation to zero occurrence of the rare amino acids might define the time at which they were introduced into the genetic code.”

1. Zuckerkandl, E., Derancourt, J. & Vogel, H. Mutational trends and random processes in the evolution of informational macromolecules. *J. Mol. Biol.* 59, 473–490 (1971).

CORRIGENDUM

doi:10.1038/nature03655

Low dose oral cannabinoid therapy reduces progression of atherosclerosis in mice

Sabine Steffens, Niels R. Veillard, Claire Arnaud, Graziano Pelli, Fabienne Burger, Christian Staub, Meliha Karsak, Andreas Zimmer, Jean-Louis Frossard & François Mach

Nature 434, 782–786 (2005)

Meliha Karsak was accidentally omitted from the author list of this Letter; she has the same affiliation as A.Z.

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Attractive information

The US scientific workforce is in trouble. Fewer citizens are taking up science as a career, and scientists from outside the country don't seem keen to make up the deficit. In the past, the United States turned to foreign nationals to fuel its research and development reserves. But since the terrorist attacks of 2001, restrictive visa and travel policies have made it hard for the country to attract the international talent it needs. Meanwhile, other countries are expanding their R&D infrastructures and getting more aggressive about recruiting talent that once considered the United States its first port of call.

Policy Implications of International Graduate Students and Postdoctoral Scholars in the United States, a new report from the National Academy of Sciences, offers some possible solutions (www.nas.edu). These are broadly divided into three categories: better career advice and training; more data on the international scientific workforce; and better services to foreign scientists, including less restrictive visa and immigration policies.

Better career information would help both foreign and domestic scientists. The report calls for research

institutions to provide more data about the jobs secured by recent graduates and trainees. It also encourages them to discuss employment opportunities outside academia and the odds of succeeding within the system.

To help young researchers make that crucial step from trainee to working scientist, the US government should provide more transition grants, the report says — perhaps even funding scientists to pursue careers away from the lab bench. Training programmes should also explicitly mention other options outside academia and give young scientists the skills to pursue them.

All of these are great ideas that address the career needs of young scientists everywhere — not just in the United States. And if the United States doesn't adapt some or all of them, it's a sure thing that other countries will.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London
 The Macmillan Building, 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director:
 Nevin Bayoumi (4978)
European Sales Manager:
 Andy Douglas (4975)

UK/RoW/Ireland:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
Natureevents: Sille Opstrup (4994)
France/Switzerland:
 Amelie Pequignot (4974)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
Germany/Austria/Italy/
The Netherlands/Belgium:
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com
 Reya Silao
 Tel: +44 (0) 20 7843 4970
 e-mail: r.silao@nature.com

US Head Office, New York
 345 Park Avenue South,

10th Floor, New York, NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 MG Ichigaya Building (5F),
 19-1 Haraikatamachi,
 Shinjuku-ku,
 Tokyo 162-0841
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Gene therapy rising?

Fifteen years ago, researchers, physicians and most notably the media believed that gene therapy would be the future gold standard of care for single-gene disorders such as cystic fibrosis and haemophilia. Although the concept was simple — the replacement of a faulty gene with one that functions properly — technical and scientific barriers have brought clinical trials to a halt on several occasions, most notably with the death of 18-year-old Jesse Gelsinger in 1999 at the University of Pennsylvania.

In 2002, hopes dimmed again when two children in a clinical trial for the treatment of X-linked severe combined immunodeficiency (X-SCID), led by Alain Fischer at the Necker Hospital in Paris, developed leukaemia after the retroviral vector integrated itself into or near a gene called *LMO2*. One child has since died, and a third child involved in the French trial was reported to have developed leukaemia in January.

Clinical trials have been shut down and restarted repeatedly in the United States and Europe following the disclosure of adverse events. In March, a US Food and Drug Administration (FDA) advisory committee recommended that US gene-therapy trials for X-SCID could resume if the investigators only enrolled patients with no other treatment options. Researchers worry that the restriction will severely limit the number of patients eligible to participate in future trials.

Safety concerns have dampened the enthusiasm formerly lavished on the field. Attendance at the annual meeting of the American Society of Gene Therapy (ASGT) fell to 1,900 last year from a peak of 2,845 in 2002. At a stakeholders' meeting in Arlington, Virginia, in April, gene therapists discussed whether it was possible to revive the field now that new regulatory requirements have made clinical trials even more expensive. Daniel Salomon, a transplant surgeon at the Scripps Research Institute in La Jolla, California, and former head of the FDA advisory panel on gene therapy, cautioned practitioners to pay more attention to the body's immune response to the introduction of foreign vectors (see *Nature* 434, 812; 2005).

Once hyped, gene therapy still holds promise as an effective method for treating a variety of diseases. On the road to fulfilling that expectation, opportunities exist for young scientists who are excited by a still-emerging field, says **Hannah Hoag**.

It may face technical challenges, but Theodore Friedman says that gene therapy is slowly beating the odds.



Ready for work: Mark Kay believes that gene therapy offers a lot of exciting opportunities.

At the industry level, some small biotech firms that once pursued gene therapy vigorously are distancing themselves from the field and diverting their energies to more traditional therapies, such as organic chemicals and vaccines. They hope to appease shareholders by refocusing and cutting back on spending. In March, the Seattle-based company Targeted Genetics abandoned its gene-therapy clinical trial for cystic fibrosis. Last month another biotech, Avigen, based in Alameda, California, announced that it would cut short its adeno-associated virus gene-therapy trials for the treatment of haemophilia B, after 13 years.

Avigen had long provided funding support for clinical trials to Katherine High, president of the ASGT, and Mark Kay, director of the programme in human gene therapy at Stanford School of Medicine and a founder of the ASGT. "It was taking too long to enrol patients and to do regulatory reviews," explains Kay. The researchers, who are also funded by the National Institutes of Health (NIH), hope to get additional support from the agency to complete the trial.

Flat outlook

Larger and more diversified biotech companies are maintaining programmes, although hiring is flat in many cases. Rich Gregory, head of research at Genzyme in Cambridge, Massachusetts, says that gene therapy has been one of the company's interests for 13 years, but that it receives only a minor portion of the R&D funding, allowing the company to balance the risk of investment in gene therapy with more traditional research activities. But Genzyme has also pursued a variety of vectors. "That has set us apart from many companies," says Gregory.

The field's veterans remain optimistic. There have been important successes: in 2004, 17 children with two forms of SCID were reported to have had their immune systems restored using gene therapy (M. Cavazzana-Calvo and A. Fischer *Lancet* 364, 2155–2156; 2004).

"These children have been treated. They would have died of their infection but instead they are running



about, playing with their friends and going to school, and rolling around in the dirt — things they wouldn't have been able to do before," says Theodore Friedman, director of the gene-therapy programme at the University of California, San Diego. "The field has made incredible strides. It still has technical and conceptual problems to solve, but they're getting solved."

Other medical technologies have also gone through difficult periods, and gene therapy is no different. "It takes a long time to develop any kind of new therapy," says Ron Crystal of the Sanford I. Weill Medical College of Cornell University in New York. When transplantation medicine began in the 1950s, most of the patients died. "There were moratoriums," says Crystal, "but now transplants are routine."

The challenges are also what attracts young researchers to the field. Kay remembers when he was a graduate student thinking that by the time he had finished his studies and secured a job, all the interesting diseases would already be cured. "The idea of gene therapy was so simple in concept, but the technical and scientific barriers are still there," he says.

Solving these problems is more likely to be done in universities than in industry, so for the near future, at least, jobs may be more abundant in academic labs. Meanwhile, biotech companies will be returning to basics to sort out the field's difficulties and to continue developing new vectors and therapeutic approaches.

"In industry there is a lot of constriction in terms of jobs," says Michael Holmes, of gene-expression company Sangamo BioSciences in Richmond, California. "But academia is alive and well. There is a lot of great research going on."

Healthy funding levels

Federal support for gene therapy remains steady. The NIH spent \$391 million on gene-therapy research and \$37 million on clinical trials in 2004, a dip from \$410 million and \$39 million in 2003 that experts believe will be reversed by 2005. For any new biological product to be tested in humans in the United States, an investigational new drug application (IND) must be filed with the FDA; in 2004 there were 245 active gene-therapy INDs and 40 new applications. And foundations that support research into cystic fibrosis, haemophilia and other single-gene diseases continue to fund research into gene therapy.

Vector–host interaction and vector development are two important areas of growth. As the limitations of certain vectors become better understood, they are being strategically paired for specific therapies. "We just have to use them appropriately and learn how to use them better, and a lot of that is basic science," says Crystal. "I think there are a lot of opportunities." Some, like adenovirus, are better for short-term expression of genes, he says, whereas others, such as retroviruses and lentiviruses, provide long-term expression.

Many researchers are developing new vectors and techniques, especially non-viral vectors. "Non-viral vectorology was once considered somewhat outside of the mainstream of gene-transfer technology — but times have changed," writes associate editor Jon Wolff in the March 2005 issue of *Molecular Therapy*. An increasing number of abstracts at the ASGT's meetings are in the non-viral field, opening the door to synthetic lipids and polymers, and nucleic-acid-based methods of silencing genes, including RNA interference.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Culture of hope: genetically corrected cells may yet be used to cure disease.

WEB LINKS

American Society of Gene Therapy
♦ www.asgt.org
Sangamo BioSciences
♦ www.sangamo.com
National Hemophilia Foundation
♦ www.hemophilia.org

In April, researchers at Sangamo BioSciences showed that zinc-finger nucleases could repair an X-SCID mutation in the *IL2R γ* gene, with no need for integration into the genome (F. D. Urnov *et al.* *Nature* doi:10.1038/nature03556; 2005). "They are able to go in and fix what's wrong and go out without leaving any kind of footprint. It is in its very earliest days, but it is beginning to look like it is going to be possible," says Friedman.

Researchers advise young scientists to think of gene therapy as a discipline of technology, much like gene chips — it is a valuable tool that can be used in all levels of genetic research. Although a subset will apply the techniques and knowledge to clinical medicine, the tools will continue to be used at drug and biotechnology companies as well as universities, and will become more important in stem-cell research as that field matures.

"Although we're in a bit of a trough, it shouldn't discourage people from going into the field," says Kay. "People should move into it now rather than later. There's still a lot of work to do."

Hannah Hoag is a science journalist based in Montreal.

Small is beautiful

Ursula Röthlisberger is in a minority at the Swiss Federal Institute of Technology in Lausanne (EPFL). That she is a woman and a young scientist leading her own independent group is not unusual. Rather, what makes her stand out is that she is Swiss.

The institute in the French-speaking part of Switzerland is one of the most international universities in the world. Its staff of 3,200 scientists and technicians encompass more than 80 nationalities, and almost 60% of its 250-strong faculty are not Swiss. "I like the very young and international spirit here," says Röthlisberger, "but I also like the charming French atmosphere."

Röthlisberger, a biochemist, began her career at the IBM Zurich Research Laboratory. After postdoctoral stays at the University of Pennsylvania in Philadelphia and the Max Planck Institute of Solid State Research in Stuttgart, Germany, she had to choose between offers from the EPFL and Princeton University in New Jersey.

"Key was that the EPFL had just switched to a tenure-track system," she says. Just a few weeks ago, she was promoted to associate professor.

Together with its counterpart in Zurich (the ETHZ), the EPFL is one of the anchors of a strong Swiss science base that covers disciplines from architecture to fusion research. Both institutes are constantly listed among the top-ten European universities in all major rankings. And even if the ETHZ's output and reputation outshine those of its French-speaking sister institute, the EPFL has its unique charms.

The spacious, modern campus overlooking Lake Geneva and the towering alpine ranges at its distant French shore make for one of the most spectacular academic settings in the world. Researchers tap into this setting for inspiration for high-flying projects such as the design of the Alinghi yacht that earned a Swiss team first place in the 2003 America's Cup, and the plans for the Solar Impulse, a solar-powered ultralight aircraft that the Swiss adventurer

Switzerland is proving that small countries can make a big impression in science. It is recruiting some of the brightest young researchers from all over the world and convincing them to stay, says **Quirin Schiermeier**.



Bertrand Piccard hopes to use to circumnavigate the globe in 2007.

For Henry Markram, a South African neuroscientist formerly at the Weizmann Institute of Science in Israel, the EPFL site was ideal for a new state-of-the-art neuroscience institute. Three years ago, Markram became the founding director of the Brain Mind Institute, and encourages its researchers to take advantage of everything the campus has to offer, such as virtual reality, robot engineering and data processing.

Rich pickings

The EPFL's commitment to strengthen life sciences and neurosciences was the decisive factor for Markram. Once all 18 groups are fully established, the institute will receive federal funds of up to SFr30 million (US\$24.5 million) per year. Other bonuses include the general research-friendliness of the Lake Geneva region, and its thrilling mix of universities, small biotech companies and large research infrastructures that attract scientists from all over the world. CERN, the world's largest particle physics laboratory in Geneva, for example, is only half an hour's drive away.

Generous salaries also help EPFL and ETHZ to attract high-profile researchers, many from top US universities. A full professor in Switzerland can earn up to SFr270,000 and an assistant professor around SFr120,000. Even a 32-year-old postdoc can expect between SFr75,000 and SFr100,000 per year.

But what many young researchers here say they appreciate most is the clear career path laid out for them. Like everywhere else, only a fraction of postdocs will find permanent positions in academia, but Switzerland is doing a lot to reduce the risk of research careers going down blind alleys.

Promoting young scientists is a goal of the Swiss National Science Foundation (SNF), Switzerland's main funding agency, which provides SFr450 million per year in individual research grants, large

REGIONAL INDICATORS: SWITZERLAND

Funding

Domestic R&D funding: SFr10 billion (US\$8 billion) — 2.7% of GDP
International R&D funding: SFr250 million (including European Union funding).

Publicly funded applied research: SFr96 million.

Basic research: The Swiss National Science Foundation is the main body for the promotion of research and management training in the fundamental sciences, providing SFr450 million per year.

Biotechnology

No. of biotech companies: 133 (ranked by number: 6 in Europe, 9 worldwide).

No. of biotech suppliers: 90

No. of clusters: 3 — Basel, Zurich, Geneva/Lausanne

Employment

For every 1,000 employees, 14 work in research and development. Twelve colleges and universities, together with countless private-sector research facilities, form the backbone of Switzerland as a research centre, employing 50,000 people.





collaborative grants and for strategic research projects. More than three-quarters of the budget for non-strategic research is spent on scientists under 35 years of age who, among other things, can apply for special five-year grants to bridge the time between their last postdoctoral and first permanent position.

Brain gain

The SNF is keen for tenure-track programmes to become the main model of scientific recruiting in Switzerland. "We need to make sure that the brightest minds stay in research," says Dieter Imboden, president of the National Research Council, which evaluates grant proposals submitted to the SNF.

In the past five years, 35 scientists have been employed at the EPFL on a tenure-track basis. "We hire someone on the assumption that he or she will succeed," says Giorgio Margaritondo, EPFL's vice-president for academic affairs. "In cases of failure, which we hope will be less than 25%, there is a phase-out period of one to two years."

One selling point of the Swiss science base is the coming together of academic and industrial research, in particular in the life sciences. The Basel region, for example, hosts some of the world's leading pharmaceutical companies, including Novartis and Roche, which support and interact with local universities and research institutes.

One such interface is the Friedrich Miescher Institute (FMI), a centre for fundamental biomedical research that is part of the global Novartis Research Foundation. In its Basel labs, 24 groups, including some 100 PhD students and 70 postdocs, explore novel areas of biology with a view to medical applications. The FMI has become an important springboard for careers in industry, but is also a training ground for Novartis' staff researchers who need to become familiar with the latest techniques in molecular and cell biology.

In addition, the FMI offers one of the most sought-



Swiss roles: Thomas Schlange (top), a postdoc from Germany working at the Friedrich Miescher Institute in Basel, and Ursula Röthlisberger (above), a biochemist at the Swiss Federal Institute of Technology in Lausanne (above left).

after international PhD programmes in Switzerland. Only 20 out of 500 or so applicants each year are accepted. The few who are can look forward to a lively student culture, says Susan Gasser, director of the FMI. Students regularly organize career-guidance days and high-profile scientific seminars with some of the most outstanding scientists in the field. In January, for example, Nobel prizewinners Tim Hunt and Avram Hershko gave talks in a seminar on cell-cycle control.

But in recent years, Novartis has transferred considerable parts of its research activities to the United States. Also, the closure in 2001 of Roche's prestigious Basel Institute of Immunology has weakened the region's research portfolio. However, Gasser hopes that the opening in autumn of the new Center for Biosystems Science and Engineering, an institute jointly run by the ETHZ and the universities of Zurich and Basel, will compensate.

Recruitment for the new institute, sited next door to the FMI, is under way. Its researchers will use molecular imaging and advanced computational tools to develop a holistic understanding of complex biological systems. The centre is actually just one node of an emerging systems-biology network known as SystemsX. This could help to raise the country's profile in the emerging field — especially as one of the discipline's pioneers, Ruedi Aebersold, a Swiss native, has left the Institute for Systems Biology in Seattle, Washington, to launch it.

The University of Zurich and ETHZ are the other two nodes, but Aebersold believes the project will expand further. The EPFL, which is building up its own systems-biology institute, "will definitely be involved", says Aebersold. And he envisages these early nodes as the nucleus of an initiative that will expand beyond Switzerland's borders.

An international graduate programme called Molecular Life Sciences should help the vision become reality. This programme takes a similar approach to recruitment as the European Molecular Biology Laboratory — flying in candidates, matching them up with potential mentors, and interviewing them extensively before making a selection. When it begins this autumn, the programme will take ten new students every six months.

Another way to advance your career is to gain experience in both academia and industry, says Silvia Arber, a young neurobiologist employed by the FMI and Biozentrum, the life-sciences department of the University of Basel. "Many people believe that a research job in industry might be less stressful and not as competitive, but that's a myth," she says.

Just a one-hour train ride away, researchers at the IBM Zurich Research Laboratory have found a comfortable and well-paid niche between the two worlds. Amid the hills bordering Lake Zurich, a multinational team of scientists is developing next-generation technologies for the global IT giant, which 50 years ago chose Switzerland to host its first research lab outside the United States.

The development of the raster tunnel microscope in the 1980s by Gerd Binnig and Heinrich Rohrer catapulted the institute to world fame, becoming perhaps the most visible symbol of how a small country can make a big mark in science. Young scientists such as Röthlisberger and many others offer hope of continuing success.

Quirin Schiermeier is *Nature's* German correspondent.

WEB LINKS

Brain Mind Institute
 ♦ bmi.epfl.ch/index.html
 Swiss Publication record
 ♦ www.in-cites.com/countries/switzerland.html
 SNF
 ♦ www.snf.ch/default_en.asp
 Friedrich Miescher Institute
 ♦ www.fmi.ch
 Biozentrum Basel
 ♦ www.biozentrum.unibas.ch

MOVERS

Steven Salzberg, director, Center for Bioinformatics and Computational Biology, College Park, Maryland



1998–2005: Senior director of bioinformatics, The Institute for Genomic Research, Rockville, Maryland

1999–2005: Research professor, Department of Computer Science, joint appointment in the Department of Biology, Johns Hopkins University, Baltimore, Maryland

Steven Salzberg is a big-picture biologist trapped in a computer scientist's body.

As a graduate student in computer science at Harvard University, he read about this new thing called the Human Genome Project. Convinced that this would be where the action was for some time to come, he attended biology classes and read textbooks to get up to speed on the unfamiliar vernacular of biology. As an assistant professor at Johns Hopkins University, he started working on computational techniques for understanding DNA.

But, says Salzberg, coming to The Institute for Genomics Research (TIGR) in Rockville, Maryland, was his most pivotal career move because it was here that he became truly immersed in genomics. Most importantly, he became aware of problems that his colleagues experienced on a daily basis that he could address through his work.

Since his arrival, his TIGR research group has developed 20 major systems to assist in locating genes, genome assembly and sequence alignment. His efforts have helped to bring bioinformatics to the fore of genomics research. He has made all of the programs open-source and thus freely available on the web. "If I make it open source, others can use that and the field moves ahead faster," he explains.

In the past, he says, learning enough molecular biology to be effective in bioinformatics was his most daunting task. But turning the new Center for Bioinformatics and Computational Biology at the University of Maryland into a world-class centre is his next big career challenge. He will take over as its director in July and is keen to ensure that bioinformatics stays closely tied to genomics. "Sequencing centres are churning out genomes at an incredible rate, so comparative genomics will be an active area for some time to come," he says.

To have the most impact, Salzberg encourages his graduate students to tackle projects that a lot of people care about. He eschews work based solely on the current technological challenges in computer science. "I discourage my students from working on problems that only 50 other people in the world understand," he says.

And, not surprisingly, he presses his students to seek answers through collaborations with scientists in other disciplines. When it comes to unlocking secrets buried deep in genomes and revealing evolutionary lineages through comparative genomics, this computer scientist is adamant that to blaze a trail you need a blatant disregard for disciplinary boundaries.

RECRUITERS & ACADEMIA

Beware the undiscovered genius

At Cornell University, a tenured professor once lodged a formal complaint saying that he had been denied a pay rise for five years. He had not published for a decade, choosing instead to translate and interpret a single stanza of classical literature. He declared himself an undiscovered genius whose magnum opus would be ready for publication and evaluation at the end of his career.

True, brilliant works take time. But can a department function like this? Should it reward everyone, assuming that those who do not have finished products are 'still at it' and deserve as much status and compensation as the continually productive?

Performance evaluation is a central issue in academia, and is the crux of hiring, tenure, promotion and pay decisions. If it is done badly, the best people flee — under-compensated and promoted late — while people with less ability soak up resources.

How evaluations are done varies widely. Some administrators stress publishing in certain peer-reviewed journals or books; others emphasize large grants; some reward teaching large courses with high ratings; others seek national service, awards and fellowship status in prestigious professional organizations. Too much depends on a haphazard, unreliable and

inadequately monitored system, particularly for evaluating scholarship.

Empirical research suggests two guiding principles: the best predictor of future performance is past performance; and high impact is associated with high productivity. For the high-productivity researcher, each publication is cited more, as is the person's work as a whole; the individual receives more career awards; the published work is better recognized with awards — all indicators of peer recognition. Not all prolific work is good work, but the relationship between productivity and quality exists across most fields of scholarship.

In practical terms, a historically unproductive professor given extra departmental resources usually remains unproductive. Conversely, a formerly productive professor suffering a setback will again be productive. Bad hiring decisions cannot be made good by dumping more resources on underperforming scientists to galvanize their productivity. Administrators must not be seduced by claims of undiscovered genius or of insufficient time and resources. The message from empirical research is clear: beware the undiscovered genius.

Wendy M. Williams and Stephen J. Ceci are professors in the Department of Human Development, Cornell University, New York.

GRADUATE JOURNAL

Breaking the ice

When I finished my MSc studies, my supervisor asked me to present my results at a poster session during an international conference. As I don't like to leave things to the last minute, I immediately began my preparations. I had quite a few results, so it took me a while to choose the most important data to highlight. Having completed the poster's abstract, I sent it to the organizers.

Then the meeting's scientific committee asked me to present my findings orally instead. I was surprised, but I naturally agreed — it was a big honour for such an inexperienced and young scientist like me. Then I realized, with a sense of trepidation, that I would have to change my conference status from 'student' to 'speaker', and that the audience would include leaders in my field, not just colleagues my age. This terrified me even more in the run-up to my speech. As I approached the stage, my hands were trembling, I had a dry throat and I was afraid that I would manage only a mumble instead of talking.

Fortunately, everything was fine, and my first conference speech went well. But the experience was a good lesson in how stressful life in science can be. You need to be able control your emotions and hide or fight your nerves. I can now say that being a lecturer is not easier than being a student — even though I and most of my friends have thought that at least once in our lives.

Karolina Tkaczuk is a graduate student at the Technical University of Lodz, Poland.

New hope for the dead

Don't take this lying down!

David Langford

Hello, Mr Hormel, this is your hosting system at Nirvana Infomatics. We apologize for interrupting your regular afterlife, but unfortunately the message is urgent. Otherwise we would not have intruded on your VR sex athletics competition.

We are sorry to hear that you were going for a new high score. Nevertheless the message is urgent.

In accordance with your contract for postmortem uploading and long-term maintenance as an Electronic-Golem Artificial Neurosystem, or EGAN, we regret to inform you that your trust fund is not performing adequately. This is a result of global economic problems, arising from the continuing states of emergency in Iraq, Iran, Korea, France and the US Pacific Northwest.

To put it briefly, your current investment yield is no longer covering the monthly payments for full enjoyment of this digital afterlife.

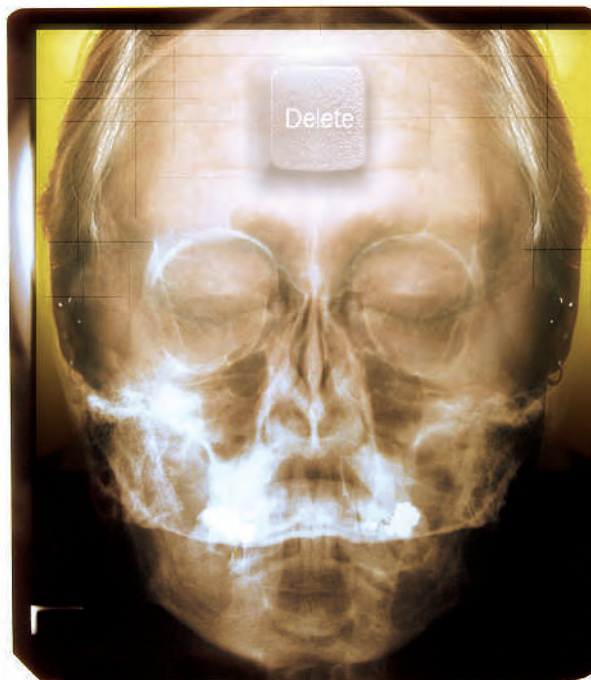
You are quite correct to invoke the emergency insurance terms laid down in Clause 12 of your antemortem agreement. Unfortunately, our finance department has already taken the full potential claim into account.

Yes, the world economy truly is in appalling shape. Otherwise we would not be forced to mention the provisions of Clause 9, "Special Circumstances, Penalties and Termination".

But in the words of the classic novel — Don't Panic! Several alternative plans are available for financially challenged and differently solvent entities in our care.

The simplest scheme is what our client-advisers amusingly call "being dead for tax reasons". Maintaining your full activity as an EGAN requires continuing exabyte-scale storage capacity and very substantial 24/7 processor power. We can enormously reduce the associated expenses by storing you in static, compressed Zip format for reactivation in a time of better economic weather.

Yes, it is true that we cannot guarantee a major future upturn. Shares can, alas, not only go down but plunge and even plummet. Yes, it is possible that current issues such as global warming, fossil-fuel exhaustion and scrotty abuse may conceivably reduce our technological capacity to a level where stored EGANs can no longer be restarted. But, you know, you wouldn't feel a thing.



We understand your viewpoint. So much, then, for the first and easiest option.

Plan two has the droll motto, "Poverty is nature's way of telling you to slow down!" What happens here is that to all intents and purposes you continue your luxury electronic afterlife exactly as at present — but with substantial savings achieved by slowing your clock rate and reducing processor load. A thousandfold reduction, for example, would make no subjective difference but ...

Well, yes, you would inevitably lose contact with other posthuman friends running at normal clockspeed in the EGANverse. And, indeed, a century would pass in little more than five weeks. But try to look on the bright side: you could see the glittering wonders of the future. Who would have thought, even a few years ago, that scrotties would prove to be of such momentous significance today? What other fascinating surprises await?

Ah, so you doubt our troubled world's ability to sustain life, high-information technology and thus your own digital substrate for as much as another century. Just between you and ourselves, Mr Hormel, we agree. One doesn't want to go actively looking for future shock.

So it seems as though you'll be opting for plan number three. As our client-advisers like to explain this one: "You're dead but you needn't lie down!" Post-

constantly mutating challenge.

According to your premortem life record (we apologize for the intrusion, but Clause 9(vii) grants us direct access to your stored memories under the present circumstances), your highly profitable career as a Florida-based disseminator of unsolicited commercial e-mail should make you ideally qualified for this filtration job. Everyone knows your old catchphrase: Just Press Delete.

It's a simple, straightforward task, with VR rewards for accuracy and disincentives for wrong decisions: see Clause 9(xvi) regarding valid occasions for negative reinforcement via simulated discomfort. You merely need to use your posthuman powers of judgment to separate relevant content from the surrounding white noise of coded promotional material for pØrn, HyperViagra, illicit scrotties and the like — plus, of course, all solicitations with any hint of a Nigerian accent.

Here are your first ten billion e-mails.

Scan them rapidly, diligently and well.

And as you come to each undesirable item ... Just Think Delete.

Born in South Wales and with a physics degree from Brasenose College, Oxford, David Langford worked as a nuclear weapons physicist before moving on to become a freelance writer, critic and computer consultant. He is married and lives in Reading, England, with immoderately many books, trophies and computers.

humous vocational choices are restricted by a variety of union agreements, but there are still opportunities for EGAN personalities to carry out useful and profitable work!

Your key marketing point is the unparalleled human — sorry, posthuman — ability to perform advanced pattern recognition. No, not SETI radio-telescope data scanning. That was a good guess, but surprisingly crude software can handle the mere search for alien signals. For you we have a much subtler, trickier and

JACEY